United States Patent  
Moyer et al.  
[19]  
[34] ENTOMOPOXVIRUS SPHEROIDIN GENE SEQUENCES  
[75] Inventors: Richard W. Moyer; Richard L. Hall, both of Gainesville, Fla.  
[21] Appl. No.: 991,867  
[22] Filed: Dec. 7, 1992  

Related U.S. Application Data  

[51] Int. Cl.  
C12N 5/10; C12N 5/11; C12N 15/39; C12N 25/62  
[52] U.S. Cl.  
435/240.2; 536/23.72; 536/24.1; 435/240.1  
[58] Field of Search  
536/23.72, 24.1; 435/320.1  

References Cited  
U.S. PATENT DOCUMENTS  
5,174,993 12/1992 Paoletti 424/199.1  

OTHER PUBLICATIONS  

Primary Examiner—Robert A. Wax  
Assistant Examiner—Eric Grimes  
Attorney, Agent, or Firm—Saliwanchik & Saliwanchik  

ABSTRACT  
The subject invention pertains to novel Entomopoxvirus (EPV) spheroidin polynucleotide sequences free from association with other viral sequences with which they are naturally associated, recombinant polynucleotide vectors containing the sequences, recombinant viruses containing the sequences, and host cells infected with the recombinant viruses are provided herein, as well as methods for use thereof in the expression of heterologous proteins in both insect and mammalian host cells.  

17 Claims, 28 Drawing Sheets  

AmEPV Hind III map:  

<table>
<thead>
<tr>
<th>A</th>
<th>C</th>
<th>B</th>
<th>J</th>
<th>F</th>
<th>G</th>
<th>D</th>
<th>H</th>
<th>E</th>
<th>10 kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' Spheroine</td>
<td>3'</td>
<td>3'</td>
<td>NPH I</td>
<td>5'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RM83 RM161 RM206 RM212

1 kb c | B | A | F | J | G | D | H | E

5000 250 bp 6088 6280 6769 8223 8457

(continued on next page.)
OTHER PUBLICATIONS


Fig. 2A

AGATCTGATG TTCTATATAT AGTACAAATT TGTATGATTA ATTGATTTT TAAAATTCAA

GATA TTA AAT ATT AGA TTC TAA ACT ATT CTT CTC ATT ATC AAT ATA ACT
Ile Asn Ser Glu Leu Ser Asn Lys Glu Asn Asp Ile Tyr Ser
1 
5

10

109

ATC ATA ATC ATT TTT TAT TTT ATT TCT ATT AAT TCT ATT ATT ACT ATT
Asp Tyr Asp Asn Lys Ile Lys Ser Cys Val Asn Met Ile Arg Asn Ser Asn
15
20
25
30

160

TTT TTT ATA CAT ATC TAT TAA TTC CAT AAA CTT TTT ATT TTT TAT ATT AAA
Lys Lys Tyr Met Asp Ile Leu Glu Met Phe Lys Lys Asn Lys Ile Asn Phe
35
40
45

211

TAT TTC TAA TGT ATT TTT AAA TTC GTC AAT ACT ATT AAT ATC ATA TCT AGA
Ile Glu Leu Thr Asn Lys Phe Glu Asp Ile Ser Asn Ile Asp Tyr Arg Ser
50
55
60
65

262

AAT AAA TAA TGC ACC TCT ATA ACT ACT AGC CAA TAA ATC ACC AAT AAA ACT
Ile Phe Leu Ala Gly Arg Tyr Ser Ser Ala Leu Leu Asp Gly Ile Phe Ser
70
75
80

313

CAT AGA ATA ATA TAA TTT TTT AAA TTC AAA TTT AGA TTT TAT GTC GAA ATA
Met Ser Tyr Tyr Leu Lys Lys Phe Glu Phe Lys Ser Lys Ile Asn Phe Tyr
85
90
95

364

AAC TRT ATA ATA AAA TAT TAT ATT AAA CAT ACC ACA ATC GGG ACT ATC
Val Ile Tyr Tyr Leu Phe Ile Ile Asn Phe Met Gly Cys Asp Pro Ser Asp
100
105
110
115

415
Fig. 2B

ATA TTG TAA TTC AAA AGT ATT AAA AAA GTA ATT TAG ATT TAG ATT TTT AAA TAT
Tyr Gln Leu Glu Phe Thr Asn Phe Phe Tyr Tyr Asn Val Asn Lys Phe Ile
120

ATC ATT TAA ATA TTC TGA TAG TAG TAC ATC ATC AAT GTA TAA ATA AGC ATA ATT AGT
Asp Asn Leu Tyr Glu Ser Leu Val Asp Ile Tyr Leu Tyr Ala Tyr Asn Thr
135

ATT AGG AGT ACT ATT GTA GTG TTT ATG GCT TTT TAT AGT CAT ATC AGA TTC
Asn Pro Thr Ser Asn Tyr His Lys His Ser Lys Ile Thr Met Asp Ser Glu
145

AAT AAA CAT ATA TTT TTT ATT TTT TAT AAG TTC TGG TAT ATA ACC ACT
Ile Phe Met Tyr Lys Lys Asn Gln Lys Ile Leu Glu Pro Ile Tyr Gly Ser
155

ACT ATT AAA AAA GTA TGC AGC TTT ATC TTT ATC AAA GTG TTT ATC TAT
Ser Asn Phe Phe Tyr Ala Ala Lys Asp Lys Asp Phe His Lys Asp Ile
160

TAC GCA ACA AGT AAA ATG ATC ATT ATA AAT TAT AGG AAA CAT AAA AAA TCT
Val Cys Cys Thr Phe His Asp Asn Tyr Ile Ile Pro Phe Met Phe Phe Arg
175

TTT TTT ATC ATT CAT TAA AAA AAA TTT TAC TCT ATC TTC AAG TTT ATA GCA
Lys Lys Asp Asn Met Leu Phe Phe Lys Val Arg Asp Glu Leu Lys Tyr Cys
190
Fig. 2E

TTC AAT ACA TTT TTA ATA ATA ATT TTA TTA TTT GGT ATT ATA GGT ATT ATT TAT 1539
Phe Asn Thr Phe Leu Ile Ile Ile Leu Leu Phe Gly Ile Ile Gly Ile Tyr
470 475 480 485

ATA TTA ACA TTT GTG TTT AAT ATA GAT TTT TTA ATA AAT AAT AAT AAA ATA 1590
Ile Leu Thr Phe Val Phe Asn Ile Asp Phe Leu Ile Asn Asn Lys Ile
490 495 500

TAT ATA TTA TCA TAT AAC GCA ACT AAT ATA AAC AAT ATA AAT AAT TTA AAT 1641
Tyr Ile Leu Ser Tyr Asn Ala Thr Asn Ile Asn Asn Leu Asn
510 515 520

TTA TAC GAT TAT TCA GAT ATT ATA TTT TTG ACA AAT TTT AAC ATA AAT AAT 1692
Leu Tyr Asp Tyr Ser Asp Ile Ile Phe Leu Thr Asn Phe Asn Ile Asn Asn
525 530 535

AAT CTT TTA GTA ACA CAA GCT AAT AAT TTA CAA GAT ATA CCA ATA TTT AAT 1743
Asn Leu Leu Val Thr Gln Ala Asn Leu Gln Asp Ile Pro Ile Phe Asn
540 545 550

GTA AAT AAT ATT ATA TCT AAT CAA TAT AAT TTT TAT TCA GCG TCT AGT AAT 1794
Val Asn Asn Ile Ile Ser Asn Gln Tyr Asn Phe Tyr Ser Ala Ser Ser Asn
555 560 565 570

AAT GTA AAT ATA TTA TTA GGA TTA AGA AAA ACA TTA AAT ATA AAT AGA AAT 1845
Asn Val Asn Ile Leu Leu Gly Leu Arg Lys Thr Leu Asn Ile Asn Arg Asn
575 580 585
Fig. 2G

ACC ATA ATA TCC AGG TCT ATA ATT ATC TTT AAA AAC TTG GGA TTG AGA TAC
Gly Tyr Tyr Gly Pro Arg Tyr Asn Asp Lys Phe Val Gln Ser Gln Ser Val
735 740 745

TTC TTC AGT TTT TAA ATT AAA ATA TCC AGG ATT TTT TTT TGA TGA
Glu Glu Thr Lys Leu Asn Asn Phe Tyr Gly Leu Asn Asn Lys Lys Ser Ser
750 755 760 765

< G3L
AGA CAT AATGATATT ATATACTTT ATAGAT ATG TCA ATA TTT ATC TAC TAT
Ser Met
770

AGA CAT AATGATATT ATATACTTT ATAGAT ATG TCA ATA TTT ATC TAC TAT
Ser Met

ATT TTC AAC AAT AGA TTT TAT ATA TAT AAA AGA ATG AAT ACT GTA CAA ATT
Ile Phe Asn Asn Arg Phe Tyr Ile Tyr Lys Arg Met Asn Thr Val Gln Ile
775 780 785 790

TAA GGT GTC ATA TTA ATA ACA ACA GCA TTA TCT TTA GTA TTT CAA TTA
Leu Val Val Ile Leu Ile Thr Ala Leu Ser Phe Leu Val Phe Gln Leu
795 800 805

TGG TAT TAT GCC GAA AAT TAC GAA TAT ATA TTA AGA TAT AAT GAT ACA TAT
Trp Tyr Tyr Ala Glu Asn Tyr Glu Tyr Ile Leu Arg Tyr Asp Thr Tyr
810 815 820 825

TCA AAT TTA CAA TTT GCG AGA AGC GCA AAT ATA AAT TTT GAT GAT TTA ACT
Ser Asn Leu Gln Phe Ala Arg Ser Ala Asn Asn Phe Asp Asp Leu Thr
830 835 840

GTT TTT GAT CCC AAC GAT AAT GTT TTT AAT GTT GAA GAA AAA TGG CGC TGT
Val Phe Asp Pro Asn Asp Asn Val Phe Asn Val Glu Glu Lys Trp Arg Cys
845 850 855

GCT TCA ACT AAT AAT AAT ATA TTT TAT GCA GTT TCA ACT TTT GGA TTT TTA
Ala Ser Thr Asn Asn Asn Ile Phe Tyr Ala Val Ser Thr Phe Gly Phe Leu
860 865 870 875
Fig. 21

TGT AAT GTA AAC GAA GCC CAA GTA TGG AAA TAT GTA AGT CTA TTA TTG CTA 3400
  Cys Asn Val Asn Glu Ala Gln Val Trp Lys Tyr Val Ser Arg Leu Leu Leu 1020
       1025      1030        1035

GAT AAT GTA TCA CAT AAT GAC GTA AAA TAT AAA TTA GCT AAT TTT AGA CTG 3451
  Asp Asn Val Ser His Asn Asp Val Lys Tyr Lys Leu Ala Asn Phe Arg Leu 1040
       1045        1050

ACT CTT AAT GGA AAA CAT TTA AAA TTA AAA GAA ATC GAT CCA CCG CTA TTT 3502
  Thr Leu Asn Gly Lys His Leu Lys Leu Lys Glu Ile Asp Gln Pro Leu Phe 1055
       1060        1065

ATT TAT TTT GTC GAT TGT GGA AAT TAT GCA TTA ATT ACT AAG GAA AAT 3553
  Ile Tyr Phe Val Asp Leu Gly Asn Tyr Gly Leu Ile Thr Lys Glu Asn 1070
       1075        1080        1085

ATT CAA AAT AAT AAT TTA CAA GTC AAC AAA GAT GCA TCA TTT ATT ACT ATA 3604
  Ile Gln Asn Asn Asn Ala Val Asp Leu Asn Leu Lys Asp Ala Ser Phe Ile Thr Ile 1090
       1095        1100

TTT CCA CAA TAT GCG TAT ATT TGT TTA GGT AGA AAA GTA TAT TTA AAT GAA 3655
  Phe Pro Gln Tyr Ala Tyr Ile Cys Leu Gly Arg Lys Val Tyr Leu Asn Glu 1105
       1110        1115        1120

AAA GTA ACT TTT GAT GTA ACT ACA GAT GCA ACT AAT ATT ACT TTA GAT TTT 3706
  Lys Val Thr Phe Asp Val Thr Thr Asp Ala Thr Asn Ile Thr Leu Asp Phe 1125
       1130        1135

AAT TCT CAA AAT AAT TTT CCA GAA GAT CAA CAA TTT ACT CAA GAA GAT TTC 3757
  Asn Lys Ser Val Asn Ile Ala Val Ser Phe Leu Asp Ile Tyr Tyr Glu Val 1140
       1145        1150

AAT AAT AAT AAT AAT AAT AAT GAA CAA AAT GAA GAT TTA AAA GAT TTA CTT 3808
  Asn Asp Asn Glu Gln Asp Leu Leu Lys Asp Leu Leu Lys Arg Tyr Gly 1155
       1160        1165        1170
Fig. 2j

GAA TTT GAA GTC TAT AAC GCA GAT ACT GGA TTA ATT TAT GCT AAA AAT CTA 3859
Glu Phe Glu Val Tyr Asn Ala Asp Thr Gly Leu Ile Tyr Ala Lys Asn Leu
1175
1180
AGT ATT AAA AAT TAT GAT ACT GTG ATT CAA GTA GAA AGG TTG CCA GTT AAT
Ser Ile Lys Asn Tyr Asp Thr Val Ile Gln Val Glu Arg Leu Pro Val Asn
1190
1195
1200
1205
TTG AAA GTT AGA GCA TAT ACT AAG GAT GAA AAT GGT CGC AAT CTA TGT TTG 3961
Leu Lys Val Arg Ala Tyr Thr Lys Asp Glu Asn Gly Arg Asn Leu Cys Leu
1210
1215
1220
RM58
ATG AAA ATA ACA TCT AGT ACA GAA GTA GAC CCC GAG TAT GCA ACT AGT AAT 4012
Met Lys Ile Thr Ser Ser Thr Glu Val Asp Pro Glu Tyr Val Thr Ser Asn
1225
1230
1235
AAT GCT TTA TTG GGT AGC CTC AGA GTA TAT AAA AAG TTT GAT AAA TCT CAT 4063
Asn Ala Leu Leu Gly Thr Leu Arg Val Tyr Lys Lys Phe Asp Lys Ser His
1240
1245
1250
1255
TTA AAA ATT GTA ATG CAT AAC AGA GGA AGT GGT AAT GTA TTT CCA TTA AGA 4114
Leu Lys Ile Val Met His Asn Arg Asp Glu Gly Ser Gly Asn Val Phe Pro Leu Arg
1260
1265
1270
TCA TTA TAT CTG GAA TTG TCT AAT GTA AAA GGA TAT CCA GTT AAA GCA TCT 4165
Ser Leu Tyr Leu Glu Leu Ser Asn Val Lys Gly Tyr Pro Val Lys Ala Ser
1275
1280
1285
1290
GAT ACT TCG AGA TTA GAT GTT GGT ATT AAC TTA AAT AAA ATT TAT GTA 4216
Asp Thr Ser Arg Leu Asp Val Gly Ile Tyr Lys Leu Asn Lys Ile Tyr Val
1295
1300
1305
GAT AAC GAC GAA AAT AAA ATT ATA TTG GAA GAA ATT GAA GCA GAA TAT AGA 4267
Asp Asn Asp Glu Asn Lys Ile Ile Leu Glu Glu Ile Glu Ala Glu Tyr Arg
1310
1315
1320
Fig. 2L

TGT ATA TTA GGA ATA AAC ATA GGT AAT TCC GTA AAT ATT AGT AGT TTG CCT
Cys Ile Leu Gly Ile Asn Ile Gly Asn Ser Val Asn Ile Ser Ser Leu Pro
1480 1485 1490

GGT TGG GTA ACA CCT CAC GAA GCT AAA ATT CTA AGA TCT GGT GTG GCT AGA
Gly Trp Val Thr Pro His Glu Ala Lys Ile Leu Arg Ser Gly Cys Ala Arg
1495 1500 1505 1510

GTT AGA GAA TTT TGT AAA TCA TTC GGT GAT CTT TCT AAT AAG AGA TTC TAT
Val Arg Glu Phe Cys Lys Ser Phe Cys Asp Leu Ser Asn Lys Arg Phe Tyr
1515 1520 1525

GCT ATG GCT AGA GAT CTC GTA AGT TTA CTA TTT ATG TGT AAC TAT GTT AAT
Ala Met Ala Arg Asp Leu Val Ser Leu Phe Met Cys Asn Tyr Val Asn
1530 1535 1540 1545

ATT GAA ATT AAC GAA GCA GTA TGC GAA TAT CCT GGA TAT GTC ATA TTA TTC
Ile Glu Ile Asn Glu Ala Val Cys Glu Tyr Pro Gly Tyr Val Ile Leu Phe
1550 1555 1560

GCA AGA GCT ATT AAA GTA ATT AAT GAT TTA TTA ATT AAC GGA GTA GAT
Ala Arg Ala Ile Lys Val Ile Asn Asp Leu Leu Ile Asn Gly Val Asp
1565 1570 1575

AAT CTA GCA GGA TAT TCA ATT TCC TTA CCT ATA CAT TAT GGA TCT ACT GAA
Asn Leu Ala Gly Tyr Ser Ile Ser Leu Pro Ile His Tyr Gly Ser Thr Glu
1580 1585 1590 1595

AAG ACT CTA CCA AAT GAA AAG TAT GGT GGT GAT AAG AAA TTT AAA TAT
Lys Thr Leu Pro Asn Glu Lys Tyr Gly Gly Val Asp Lys Phe Lys Tyr
1600 1605 1610

CTA TTC TTA AAG AAT AAA CTA AAA GAT TTA ATG CGT GAT GCT GAT TTT GTC
Leu Phe Leu Lys Asn Lys Leu Lys Asp Leu Met Arg Asp Ala Asp Phe Val
1615 1620 1625 1630
Fig. 2M

CAA CCT CCA TTA TAT ATT TCT ACT TAC TTT AGA ACT TTA TGG GAT GCT CCA 5236
Gln Pro Pro Leu Tyr Ile Ser Thr Tyr Phe Arg Thr Leu Leu Asp Ala Pro 1635

CCA ACT GAT AAT TAT GAA AAA TAT TTG GTT GAT TCG TCC GTA CAA TCA CAA 5287
Pro Thr Asp Asn Tyr Glu Lys Tyr Leu Val Asp Ser Ser Val Gln Ser Gln 1650

GAT GTT CTA CAG GGT CTG TTG AAT ACA TGT AAT ACT ATT GAT ACT AAT GCT 5338
Asp Val Leu Gln Gly Leu Leu Asn Thr Cys Asn Thr Ile Asp Thr Asn Ala 1665

AGA GTT GCA TCA AGT GTT ATT GGA TAT GTT TAT GAA CCA TGC GGA ACA TCA 5389
Arg Val Ala Ser Ser Val Ile Gly Tyr Val Tyr Glu Pro Cys Gly Thr Ser 1685

GAA CAT AAA ATT GGT TCA GAA GCA TTG TGT AAA ATG GCT AAA GAA GCA TCT 5440
Glu His Lys Ile Gly Ser Glu Ala Leu Cys Lys Met Ala Lys Glu Ala Ser 1700

AGA TTA GGA AAT CTA GGT TTA GTA AAT CGT ATT AAT GAA AGT AAT TAC AAC 5491
Arg Leu Gly Asn Leu Gly Leu Val Asn Arg Ile Asn Glu Ser Asn Tyr Asn 1720

AAA TGT AAT AAA TAT GGT TAT AGA GGA GTA TAC GAA AAT AAC AAA CTA AAA 5542
Lys Cys Asn Lys Tyr Gly Tyr Arg Gly Val Tyr Glu Asn Asn Lys Leu Lys 1735

ACA AAA TAT TAT AGA GAA ATA TTG GAT TGT AAT CCT AAT AAT AAT GAA 5593
Thr Lys Tyr Tyr Arg Glu Ile Phe Asp Cys Asn Pro Asn Asn Asn Glu 1750

TTA ATA TCC AGA TAT GGA TAT AGA ATA ATG GAT TTA CAT AAA ATT GGA GAA 5644
Leu Ile Ser Arg Tyr Gly Tyr Arg Ile Met Asp Leu His Lys Ile Gly Glu 1770
Fig. 2O

ATATTATATA ATATTAACCTT ACAAGTTATA AAAAACTTA AAATGATTTT TTAARATGAT 6161
ATTATCGATA GTTGATGATAA TGTGCTCTTT TATTTATTA ATGCCGATGA TTATATATT 6221
ATCTTTTAGA TATATTTAAT ATTAAATTATA AATCGACTGA CRAATAATATT TATTC CTA 6279
TTC ATA ATA ATC ATC TGC TAT ATA TAT TAA TGT ATC ATT CTC TAT TAT AAA 6330
Glu Tyr Tyr Asp Asp Ala Ile Tyr Ile Leu Thr Asp Asn Glu Ile Ile Ile Phe 1935 1940
1945
TAT AGG TAT ATT GTC TTT ATC AAT CAT TAA TTT TGC TAC AGC TGT ATT ATC 6381
Ile Pro Ile Asn Asp Lys Asp Ile Met Leu Lys Ala Val Ala Thr Asn Asp 1950 1955 1960 1965
TTT ATA TAC TAT ATT TGT GTC TTT GTT TAA TAA ACC TTT TAA TAT AGT GGC 6432
Lys Tyr Val Ile Asn Thr Asp Lys Asn Leu Leu Gly Lys Leu Ile Thr Ala 1970 1975 1980
TCT ATC ATA ATC TTT ACA ATA TGA TAT GGG ATA TAA TTT TAT ATT AAT AAT 6483
Arg Asp Tyr Asp Lys Cys Tyr Ser Ile Pro Tyr Leu Lys Ile Asn Ile Ile 1985 1990 1995
AAC ATT AGA TAC GTT CAT TTT CAT TCT AGT TTT ACG TAT GTC AAA 6534
Val Asn Ser Val Asn Met Glu Lys Met Arg Thr Lys Arg Ile Thr Asp Phe 2000 2005 2010 2015
AAT TAT TTC ATT TTC TGC TGG TTC TAT ATA TTT ATA TGT GTG AGT AAT AGA 6585
Ile Ile Glu Asn Glu Ala Pro Glu Ile Tyr Lys Tyr Thr Asn His Ile Ser 2020 2025 2030
TTC GAT AGA TGA TTA TAA ATC AAA TAT AAC ATT TAT TTT ACC TGG 6636
Glu Ile Ser Ser Ser Lys Leu Leu Asp Phe Ile Val Asn Ile Lys Gly Gln 2035 2040 2045 2050
Fig. 2P

TTT ATC TTT TAT AAT ATC TAA TAT TTC TTT ATC TAC AGA TTT TCT GTT GTT 6687
Lys Asp Lys Ile Ile Asp Leu Ile Glu Lys Asp Val Ser Lys Arg Asn Asn 2055 2060 2065

GGT ATA TGA TAT TAA AAA ATG AAC GTT AAC ATA TCT ATA TTC TTG TGG TAA 6738
Thr Tyr Ser Ile Leu Phe His Val Asn Val Tyr Arg Tyr Glu Gln Pro Leu 2070 2075 2080

< G6L

ATC TTT ATG AGA ATT TAA TCT TAT AGA TCT 6768
Asp Lys His Ser Asn Leu Arg Ile Ser Arg 2085 2090 2094
Fig. 3B

AGA ATT AAA TAT ATT TCT GTT AAA GTC ACA ATT TAA TCC AGC AAC AAT AAC 461
Ser Asn Phe Ile Asn Arg Asn Phe Asp Cys Asn Leu Gly Ala Val Ile Val
125 130 135

TTT TTT TTT ATT ATT AGC CAT TTT ATC ACA AAA TTG TTC TAA ATC ATT TTC 512
Lys Lys Lys Asn Asn Ala Met Lys Asp Cys Phe Gln Glu Leu Asp Asn Glu
145 150 155

TTC AAA AAC TTG ACA CTC ATC TAT GCC AAT AAT ATC ATA ATT ATC TAC GAT 563
Glu Phe Phe Gln Cys Glu Asp Ile Gly Ile Ile Asp Tyr Asn Asp Val Ile
160 165 170 175

ATT GAT TTC ATT AAT TAA ATT ATT TGT TTT ATT GTA TAA ATA TTC TTT ATT 614
Asn Ile Glu Asn Ile Leu Asn Asn Thr Lys Ile Tyr Leu Tyr Glu Lys Asn
180 185 190

TAA TAT ATT TCC GTC ATG ATT TAT TAT ATT TCT ATT ATC Leu Ile Asn Gly Asp His Asn Ile Ile Asn Lys Asn Ile Phe Arg Asn Asp
195 200 205

TAT ATT ATG AGT TAT AAC TAC ACA TTT TTG ATT AGA TAA AAT ATA TCT ATT 716
Ile Asn His Thr Ile Ile Val Cys Lys Gln Asn Ser Leu Ile Tyr Arg Asn
210 215 220 225

AAT TTT TCG CAT CAA TTC TGT TGT TTT GCC AGA AAA CAT AGG ACC AAT TAT 767
Ile Lys Arg Met Leu Glu Thr Thr Lys Gly Ser Phe Met Pro Gly Ile Ile
230 235 240

< ORF Q2
TAA TTC TAT CGA CAT TTTTTTTAT TATTTGATAT ATTTTTCAA AAAAAAATTA
Leu Glu Ile Ser Met
245

ORF Q3 >
ATCAATGAAA AAAAAATAAA ATTATCAA ATG GAT TTA CTA AAT TCT GAT ATA ATT 878
Met Asp Leu Leu Asn Ser Asp Ile Ile
250 255
Fig. 3D

TTT ATA GAA AAA TTA ATT AAT TTA AAA AAA TTA GAT ATA TCT TTC AAT GTT 1388
Phe Ile Glu Lys Leu Ile Asn Leu Lys Lys Leu Asp Ile Ser Phe Asn Val
415

AAA AAA AAT ATA CAT TTG ATA AAA TTT CCA AAA AGT ATA ACT CAT TTA 1439
Lys Lys Asn Asn Ile His Leu Ile Lys Phe Pro Lys Ser Ile Thr His Leu
430

TGT GAT TAT CAA TCA TAT AAA GAA AAT TAT AAT TAT TTA AAA AAT TTA TCA 1490
Cys Asp Tyr Glu Ser Tyr Lys Glu Asn Tyr Asn Tyr Leu Lys Asn Leu Ser
445

450

455

460

AAT ATA ATT GAA TAT GAA TTC 1511
Asn Ile Ile Glu Tyr Glu Phe
465
Fig. 4

Hind III map:

AmEPV

C  B  A

F  |  |  G  D  H  E

B  B  B

0.5 kb

4.51 kb

931

H

B

0

1.4 kb

1.88 kb

307 bp  80 bp

G6L

G5R

G4R

G2L

G1L

0.7 kb

0.5 kb

1.88 kb

Spheroidin

Vaccinia

NTPase I

Capripoxvirus

HM3

Vaccinia

ORF17

Amino Acid Homologies

5,476,781
AmEPV Spheroidin, 3009 bp

5' - 12 - 9 - 8 - 6 - 3 - 4 - 11/5 - 13 - 2 - 57 - 14 - 10 - 250 bp - 3'
Fig. 6A

**AmEPV**
AAAAAGTTTGATAAATCTCAATTAAATTATTGATAATGCATAACAGAGGAAGT

**CbEPV**
1AAAAAGTTTGATAAATCTCAATTAAATTATTGATAATGCATATAGAGGAAGT

**CfEPV**
1AAAAAGTTTGATAAATCTCAATTAAATTATTGATAATGCATATAGAGGAAGC

**AmEPV**
GGTAATGTATTTCCATTAAAGATCATATTATATCTGGAAATTGTCTAATGTA

**CbEPV**
51GGTAATGTATTTCCATTAAAGATCATATTATATCTGGAAATTGTGAACGTC

**CfEPV**
51GGTAATGTATTTCCATTAAAGATCATATTATATCTGGAAATTGTGAACGTC

**AmEPV**
AGGATATCCGTTAAAGACATCTGATACTTGCAGATATTGTTGGTATTT

**CbEPV**
101AGGTTATCCGTTAAAGACATCTGATACTTGCAGATATTGTTGGTATTT

**CfEPV**
101AGGTTATCCGTTAAAGACATCTGATACTTGCAGATATTGTTGGTATTT

**AmEPV**
ACAAATTAAATAATAATTATGATAGTACACGACAGAAATAATTATA

**CbEPV**
151ATAAACTAAAATAATAATATATTGATAGTACACGACAGAAATAATTATA

**CfEPV**
151ATAAACTAAAATAATAATATATTGATAGTACACGACAGAAATAATTATA

**AmEPV**
GAAAGAAATGAGACAGAATATAGATGCGGAAGACA

**CbEPV**
201GAAAGAAATGAGACAGAATATAGATGCGGAAGACA 235

**CfEPV**
201GAAAGAAATGAGACAGAATATAGATGCGGAAGACA 235
Fig. 6B

<table>
<thead>
<tr>
<th></th>
<th>AmEPV</th>
<th>CbEPV</th>
<th>CfEPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>323</td>
<td>KFDKSHLKIVMHRGSGNVFLRSVLYLELNSVKGYPVKASDTSRDLVGYY</td>
<td>KFDKSHLKIVMHRGSGNVFLRSVLYLELNSVKGYPVKASDTSRDLVGYY</td>
<td>KFDKSHLKIVMHRGSGNVFLRSVLYLELNSVKGYPVKASDTSRDLVGYY</td>
</tr>
<tr>
<td>373</td>
<td>KLNKICYVDNENKIIIEEIEAEYRCGR</td>
<td>KLNKICYVDNENKIIIEEIEETDYRCGR 77</td>
<td>KLNKICYVDNENKIIIEEIEETDYRCGR 77</td>
</tr>
</tbody>
</table>

Fig. 6C

<table>
<thead>
<tr>
<th></th>
<th>AmEPV</th>
<th>CbEPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>211</td>
<td>KFKYLFLKKNK</td>
<td>KFKYLFLKKNK 10</td>
</tr>
<tr>
<td>682</td>
<td>KSVNIAVSDL</td>
<td>KSVNIAVSDL 11</td>
</tr>
<tr>
<td>726</td>
<td>KYLVDSSVQSQ</td>
<td>KYLVDSSVQSQ 11</td>
</tr>
</tbody>
</table>
Fig. 7B

FESINNTVPTASMCLIK
AATTCAGAATATATATAGTTACTGGAGTTACCGGACATACATAATATTTT

NEFNAKILKXYIPT
ATTATTTGCAATTGCTATTATATTTATTATTAAAAAAATAGGATCAA

FRNEDKSTRSIILKH
AATTTCTTCTGTTATTATCTTTTTAAACAGTTCTTGATATATTATTTGAA

VEDIIIILRSKKNLSEE
ACTTCGTCTTTATTATTAGTAACTCTACCTTTTTTTTAATTAGAAGAACTTTT

ISRYINNFKDLSSSSD
TATAGATCTATATATATTAAATTTATAACTTAACTGATGACGAAATCAT

YYIFSTDISIYSRIT
AATATATAAATTTATATTACGTTCTGATATATGACTTTATGATTA

NLWPDIYLSKKIFIILI
TTTACCAAGGATCTAGTATAATGACTTTTTTTATAATAATTAAAATTT

WRPFLKIKIYIYVAT
CCATCTTGGAAATAATCTTTTTATATTTATTAAATACACAGCAGTAA

LTGKMGTDFWFLLMNSLNN
ATGTTTTTCCCCATACCAGTATCCAAAAATAATAACATACACTTTCAAATTT

KLGIFIRSVFYQYDQLT
TTTAACTCTAGAATTTCTACTTCAAAATATTGATAATCTTGTAATGT

IETNTINNMHKMLPHQ
AATTTGAGTATTTTGAATATTATTACAATTTTATTAGGCAAATAGTGG

TKDLAYNHKGVISDLA
TTTTATCGTCAATTATTATATGTTTACCAACAATAGATCTCAATGCA

< AmEPV NPHI
FM

AACATTTAGTTATATAAAAATAATATTTTATAACTTTAGATGTTCA

TTAATTTTATGTCTGATGTTGGAATCCAAAGATATGGATAATATC

TATATCATTATTTTTTTTGAATCTATGCTATCAATCGCAAAATTTAT

CCAGTATAATTATTCCAGTATTTTCCGCATATAACAAACATACATA

ATGTTGAGTGGTTGGTCTCGGAGTGAAGACGCTACTTTT

8457
ENTOMOPOXVIRUS SPHEROIDIN GENE SEQUENCES

This invention was made with Government support under Grant No. RO1 AI5722-12 awarded by the National Institutes of Health and NIH Training Grant T32 AI-07110. The Government has certain rights in this invention.

CROSS-REFERENCE TO A RELATED APPLICATION

This is a continuation-in-part of PCT application WO92/14818, which is a continuation-in-part of U.S. application Ser. No. 07/827,685, filed Jan. 30, 1992, now abandoned, which is a continuation-in-part of U.S. application Ser. No. 07/657,584, filed Feb. 19, 1991, now abandoned.

FIELD OF THE INVENTION

This invention relates generally to the field of recombinantly-produced proteins and specifically to novel, recombinant Entomopoxvirus genes, proteins, protein regulatory sequences and their uses in expressing heterologous genes in transformed hosts.

BACKGROUND OF THE INVENTION

Poxviruses are taxonomically classified into the family Chordopoxvirinae, whose members infect vertebrate hosts, e.g., the Orthopoxvirus vaccinia, or into the family Entomopoxvirinae. Very little is known about members of the Entomopoxvirinae family other than the insect host range of individual members. One species of Entomopoxvirus (EPV) is the Amsacta moorei Entomopoxvirus (AmEPV), which was first isolated from larvae of the red hairy caterpillar Amsacta moorei (Roberts and Granados [1968] J. Invertebr. Pathol. 12:141-143). AmEPV is the type species of genus B of EPVs and is one of three known EPVs which will replicate in cultured insect cells (R. R. Granados et al. [1976] "Replication of Amsacta moorei Entomopoxvirus and Autographa californica Nuclear Polyhedrosis Virus in Hemoocyte Cell Lines from Exigene aerea," in Invertebrate Tissue Culture Applications in Medicine, Biology, and Agriculture, New York, Academic Press, New York, pp. 379-389; T. Hukuhara et al. [1990] J. Invertebr. Pathol. 56:222-232; and Sato, T. [1989] "Establishment of Eight Cell Lines from Neonate Larvae of Torticids (Lepidoptera), and Their Several Characteristics Including Susceptibility to Insect Viruses," in Invertebrate Cell Systems Applications, J. Mitsuhashi (ed.), Vol. II, CRC Press, Inc., Boca Raton, Fla., pp. 187-198.


The viral replication cycle of AmEPV resembles that of other poxviruses except for the appearance of occluded virus late in infection. For AmEPV, once a cell is infected, both occluded and extracellular virus particles are generated. The mature occlusion body particle, which is responsible for environmentally protecting the virion during infection, consists of virus embedded within a crystalline matrix consisting primarily of a single protein, spheroidin. Spheroidin, the major structural protein of AmEPV, has been reported to be 110 kDa in molecular weight and to consist of a high percentage of charged and sulfur-containing amino acids (Langridge and Roberts [1982] J. Invertebr. Pathol. 39:346-353).

Another insect virus is the baculovirus. Like baculoviruses, a characteristic feature of entomopoxviruses is the amalgamation of virions within environmentally stable occlusion bodies. It is this occluded form of the virus that is primarily responsible for dissemination to other insects. While the major protein (polychelin) of baculovirus occlusions is quite similar between viruses, it has been reported that the major occlusion body protein (spheroidin) of two group B entomopoxviruses, Amsacta moorei (AmEPV) and Choristoneura biennis (CbEPV) is quite different both in terms of amino acid sequence and coding capacity of the corresponding spheroidin genes (115 and 47 kDa for AmEPV and CbEPV, respectively).


We have investigated the spheroidin genes of Choristoneura biennis, Choristoneura fumiferana, and Amsacta moorei viruses. Our results indicate, in contrast to published results, that the known Choristoneura EPV spheroidin assignment is likely incorrect and that the Choristoneura spheroidin is instead a highly conserved homolog of the AmEPV spheroidin.

The use of viruses and virus proteins in eukaryotic host-vector systems has been the subject of a considerable amount of investigation and speculation. Many existing viral vector systems suffer from significant disadvantages and limitations which diminish their utility. For example, a number of eukaryotic viral vectors are either tumorigenic or oncogenic in mammalian systems, creating the potential for serious health and safety problems associated with resultant gene products and accidental infections. Further, in some eukaryotic host-viral vector systems, the gene product itself exhibits antiviral activity, thereby decreasing the yield of that protein.

In the case of simple viruses, the amount of exogenous DNA which can be packaged into a simple virus is limited. This limitation becomes a particularly acute problem when the genes used are eukaryotic. Because eukaryotic genes usually contain intervening sequences, they are too large to fit into simple viruses. Further, because they have many
5,476,781

restriction sites, it is more difficult to insert exogenous DNA into complex viruses at specific locations.


Additionally, studies with vaccinia virus have demonstrated that poxviruses have several advantageous features as vaccine vectors. These include the ability of poxvirus-based vaccines to stimulate both cell-mediated and humoral immunity, minimal cost to mass produce vaccine and the stability of the lyophilized vaccine without refrigeration, ease of administration under non-sterile condition, and the ability to insert at least 25,000 base pairs of foreign DNA into an infectious recombinant, thereby permitting the simultaneous expression of many antigens from one recombinant.

There exists a need in the art for additional viral compositions and methods for use in expressing heterologous genes in selected host cells, and in performing other research and production techniques associated therewith.

**BRIEF SUMMARY OF THE INVENTION**

This invention pertains to novel vectors useful for producing proteins via the expression of a heterologous gene in a novel expression system. More particularly, this invention relates to methods for incorporating a selected heterologous gene (also referred to as exogenous DNA) into a poxvirus genome to produce a recombinant expression vector capable of expression of the selected gene in a host cell.

The expression systems described herein utilize novel structural and/or regulatory DNA elements from Entomopoxvirus genomes. For example, according to the subject invention, the entomopoxvirus spheroidin gene and/or the thymidine kinase gene can be used as the location for the insertion of exogenous DNA. These Entomopoxvirus genes have been discovered to be attractive sites for insertion of heterologous genes because it is possible to transfer the strongly expressed spheroidin gene, or the thymidine kinase gene, as an expression cassette, not only in insect cells, but for use in vertebrate poxviruses such as vaccinia and swinepox virus.

Another aspect of the subject invention pertains to the use of the entomopoxvirus spheroidin or thymidine kinase gene regulatory sequences in other virus vector systems to enhance the performance of those systems. Thus, the subject invention further pertains to the use of regulatory elements from entomopoxvirus to construct novel chimeric vaccines and expression systems which are functional across genera of mammalian poxviruses.

As one aspect, the invention provides novel Entomopoxvirus polynucleotide sequences, free from other viral sequences with which the Entomopoxvirus sequences are associated in nature. Specifically, the subject invention provides nucleotide sequences of Entomopoxvirus spheroidin and thymidine kinase genes, including flanking sequences and regulatory sequences. In particular embodiments, the spheroidin DNA sequence is that which occurs in the Chrionista biennis, Chrionista funifera, or Amsacta moorei Entomopoxviruses. Also specifically exemplified is the Amsacta moorei Entomopoxvirus thymidine kinase nucleotide sequence. As explained more fully herein, fragments and variants of the exemplified sequences are within the scope of the subject invention. Fragments and variants can be any sequence having substantial homology with the exemplified sequences so long as the fragment or variant retains the utility of the exemplified sequence. One specific type of variant pertains to spheroidin or tk genes from Entomopoxviruses other than those specifically exemplified herein. As described herein, for example, the current inventors have discovered that the spheroidin genes are highly conserved among different species of Entomopoxivirus. Specifically exemplified herein are three different Entomopoxvirus spheroidin genes having a high degree of homology. Other such spheroidin variants or tk variants from other Entomopoxviruses could be readily located and used by the ordinarily skilled artisan having the benefit of the subject application.

As another aspect, the present invention provides recombinant polynucleotide sequences comprising a sequence encoding an Entomopoxvirus spheroidin protein and its regulatory sequences, or a variant or fragment of the spheroidin sequence, linked to a second polynucleotide sequence encoding a heterologous gene. One embodiment of such a polynucleotide sequence provides a spheroidin promoter sequence operably linked to a heterologous gene to direct the expression of the heterologous gene in a selected host cell. Another embodiment provides a sequence encoding a spheroidin protein linked to the heterologous gene in a manner permitting expression of a fusion protein. Still another embodiment provides the heterologous gene inserted into a site in a spheroidin gene so that the heterologous gene is flanked on both termini by spheroidin sequences.

Yet a further aspect of the invention provides a recombinant polynucleotide sequence encoding an Entomopoxvirus tk gene and/or its regulatory sequences, or a variant or fragment thereof, linked to a second polynucleotide sequence encoding a heterologous gene. One embodiment of this polynucleotide sequence provides the tk promoter sequence operably linked to the heterologous gene to direct the expression of the heterologous gene in a selected host cell. Another embodiment provides the sequence encoding the tk protein linked to the heterologous gene in a manner permitting expression of a fusion protein. Still another embodiment provides the heterologous gene inserted into a site in the tk gene so that the heterologous gene is flanked on both termini by tk sequences.

Another aspect of the invention pertains to Entomopoxvirus spheroidin polypeptides, fragments thereof, or analogs thereof, optionally fused to a heterologous protein or peptide. Also provided is an Entomopoxvirus tk polypeptide, fragments thereof, or analogs thereof, optionally linked to a heterologous protein or peptide.

Yet another aspect of the invention is provided by recombinant polynucleotide molecules which comprise one or more of the polynucleotide sequences described above. This molecule may be an expression vector or shuttle vector. The molecule may also contain viral sequences originating from a virus other than the Entomopoxvirus which contributed a spheroidin or tk polynucleotide sequence, e.g., vaccinia.

In another aspect, the present invention provides a recombinant virus comprising a polynucleotide sequence as described above. Also provided are host cells infected with one or more of the described recombinant viruses.
The present invention also provides a method for producing a selected polypeptide comprising culturing a selected host cell infected with a recombinant virus, as described above, and recovering said polypeptide from the culture medium.

As a final aspect, the invention provides a method for screening recombinant host cells for insertion of heterologous genes comprising infecting the cells with a recombinant virus containing a polynucleotide molecule comprising the selected heterologous gene sequence linked to an incomplete spheroidin or tk polynucleotide sequence or inserted into and interrupting the coding sequences thereof so that the heterologous gene is flanked at each termini by an Entomopoxvirus spheroidin or tk polynucleotide sequence. The absence of occlusion bodies formed by the expression of a spheroidin protein in the spheroidin-containing cell indicates the integration of the heterologous gene. Alternatively, the absence of the thymidine kinase function, i.e., resistance to methotrexate or a nucleotide analogue of methotrexate, formed by the integration of the inactive thymidine kinase sequence indicates the insertion of the heterologous gene.

Other aspects and advantages of the present invention are described further in the following detailed description of embodiments of the present invention.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO. 1 is the DNA sequence of the Amsacta moorei Entomopoxivirus spheroidin gene and flanking sequences.

SEQ ID NO. 2 is the amino acid sequence encoded by the G1L ORF.

SEQ ID NO. 3 is the amino acid sequence encoded by the G2R ORF.

SEQ ID NO. 4 is the amino acid sequence encoded by the G3L ORF.

SEQ ID NO. 5 is the amino acid sequence encoded by the G4R ORF.

SEQ ID NO. 6 is the deduced amino acid sequence of the spheroidin protein.

SEQ ID NO. 7 is the amino acid sequence encoded by the G6L ORF.

SEQ ID NO. 8 is the DNA sequence of the Amsacta moorei Entomopoxivirus thymidine kinase (tk) gene and flanking sequences.

SEQ ID NO. 9 is a small peptide of 66 amino acids potentially encoded by ORF Q1.

SEQ ID NO. 10 is the amino acid sequence encoded by the Q3 ORF.

SEQ ID NO. 11 is the deduced amino acid sequence of the tk protein.

SEQ ID NO. 12 is the synthetic oligonucleotide designated RMS8.

SEQ ID NO. 13 is the synthetic oligonucleotide designated RM82.

SEQ ID NO. 14 is the synthetic oligonucleotide designated RM83.

SEQ ID NO. 15 is the synthetic oligonucleotide designated RM92.

SEQ ID NO. 16 is the synthetic oligonucleotide designated RM118.

SEQ ID NO. 17 is the synthetic oligonucleotide designated RM165.

SEQ ID NO. 18 is the synthetic oligonucleotide designated RM03.

SEQ ID NO. 19 is the synthetic oligonucleotide designated RM04.

SEQ ID NO. 20 is the synthetic oligonucleotide designated RM129.

SEQ ID NO. 21 is the spheroidin gene coding sequence spanning nucleotides #3080 through #6091 of SEQ ID NO. 1.

SEQ ID NO. 22 is a fragment of the spheroidin gene nucleotides #2781 through 3191 of SEQ ID NO. 1 which is likely to contain the promoter sequence.

SEQ ID NO. 23 is the G2R ORF.

SEQ ID NO. 24 is the G4R ORF.

SEQ ID NO. 25 is the G1L ORF.

SEQ ID NO. 26 is the G3L ORF.

SEQ ID NO. 27 is the G6L ORF.

SEQ ID NO. 28 is the tk gene coding sequence spanning nucleotides #234 through #782 of SEQ ID NO. 8.

SEQ ID NO. 29 is a fragment of the tk gene nucleotides #783 through #851 of SEQ ID NO. 8.

SEQ ID NO. 30 is a fragment spanning nucleotides #750 through #890 of SEQ ID NO. 8 which is likely to contain the promoter sequence.

SEQ ID NO. 31 is the Q1 ORF.

SEQ ID NO. 32 is the Q3 ORF.

SEQ ID NO. 33 is a fragment included within the sequence spanning nucleotides #2274 through #6182 of SEQ ID NO. 1 containing the entire spheroidin open reading frame and some flanking sequences.

SEQ ID NO. 34 is a polypeptide cleavage product according to the subject invention.

SEQ ID NO. 35 is a polypeptide cleavage product according to the subject invention.

SEQ ID NO. 36 is a polypeptide cleavage product according to the subject invention.

SEQ ID NO. 37 is the peptide sequence encoded by the RMS8 probe.

SEQ ID NO. 38 is a nucleotide fragment spanning nucleotides #4883 through #4957 of SEQ ID NO. 1.

SEQ ID NO. 39 is a nucleotide fragment spanning nucleotides #3962 through #4012 of SEQ ID NO. 1.

SEQ ID NO. 40 is a nucleotide fragment spanning nucleotides #4628 through #4651 of SEQ ID NO. 1.

SEQ ID NO. 41 is the AmEPV NPHI nucleotide sequence shown in FIG. 7.

SEQ ID NO. 42 is the AmEPV NPHI amino acid sequence from FIG. 7. This sequence is in the order opposite that shown in the Figure.

SEQ ID NO. 43 is the CbEPV nucleotide sequence shown in part A of FIG. 6.

SEQ ID NO. 44 is the CbEPV amino acid sequence shown in part B of FIG. 6.

SEQ ID NO. 45 is the CiEPV nucleotide sequence shown in part A of FIG. 6.

SEQ ID NO. 46 is the CiEPV amino acid sequence shown in part B of FIG. 6.

SEQ ID NO. 47 is the CbEPV amino acid sequence corresponding to amino acids 211 to 221 of AmEPV.

SEQ ID NO. 48 is the CbEPV amino acid sequence corresponding to amino acids 682 to 691 of AmEPV.
SEQ ID NO. 49 is the CbEPV amino acid sequence corresponding to amino acids 726-736 of AmEPV.
SEQ ID NO. 50 is the sequence of RM206.
SEQ ID NO. 51 is the sequence of RM212.
SEQ ID NO. 52 is the sequence of RM58.
SEQ ID NO. 53 is the sequence of RM75.
SEQ ID NO. 54 is the sequence of RM76.
SEQ ID NO. 55 is the sequence of RM78.
SEQ ID NO. 56 is the sequence of RM79.
SEQ ID NO. 57 is the sequence of RM82.
SEQ ID NO. 58 is the sequence of RM83.
SEQ ID NO. 59 is the sequence of RM87.
SEQ ID NO. 60 is the sequence of RM91.
SEQ ID NO. 61 is the sequence of RM92.
SEQ ID NO. 62 is the sequence of RM93.
SEQ ID NO. 63 is the sequence of RM95.
SEQ ID NO. 64 is the sequence of RM118.
SEQ ID NO. 65 is the sequence of RM169.
SEQ ID NO. 66 is the sequence of RM170.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a physical map of AmEPV illustrating restriction fragments thereof and showing the spheroidin gene just to the 3' end of site #29 in the HindIII-G fragment.

FIG. 2 provides the AmEPV DNA sequence of the Amsacta moorei Entomopoxvirus spheroidin gene and flanking sequences (SEQ ID NO. 1), the deduced amino acid sequences of the spheroidin protein (SEQ ID NO. 6), and five additional open reading frames (ORFs). The complete sequence of the G6 ORF is not shown in this figure but is provided in SEQ ID NO. 1.

FIG. 3 provides the DNA sequence of the Amsacta moorei Entomopoxovirus thymidine kinase (tk) gene and flanking sequences (SEQ ID NO. 8), the deduced amino acid sequences of the tk protein (SEQ ID NO. 11), and two additional ORFs.

FIG. 4 is a schematic map of an AmEPV fragment illustrating the orientation of the spheroidin ORF on the physical map and indicating homologies.

FIG. 5 shows the location of the AmEPV spheroidin specific oligonucleotide primers used in PCR to identify an AmEPV spheroidin-like gene in Choristoneura EPV DNAs. The arrowheads over the numbers in the diagram represent oligonucleotide primers. The arrowhead box shows the approximate starting point of the primer, and the 5' to 3' direction is shown by the direction of the arrowhead. Primers from the upper line are paired with various primers from the lower line as shown in Table 2 for PCR reactions. The sequences of the primers are shown in Table 2.

FIG. 6 shows sequences of corresponding regions of the spheroidin-like gene of CbEPV and CbEPV and the spheroidin of AmEPV. Panel A shows the AmEPV spheroidin sequence and PCR product sequences derived from CbEPV and CbEPV DNA using RM58 as the sequencing primer for the 1 kb PCR products resulting from the AmEPV spheroidin specific primer pair RM58-RM118 (Table 2) and either CbEPV or CbEPV DNA. The alignment of the Choristoneura EPV sequences with bases 4044-4278 of AmEPV spheroidin is shown. The predicted amino acid sequences from the sequences shown in Panel A are shown in Panel B. Identity is shown by vertical lines and two degrees of conserved changes are indicated by periods and colons.

Panel C shows CbEPV spheroidin amino acid sequences derived from protein microsequencing of 3 lys-endo-protease fragments. Corresponding AmEPV spheroidin sequence and amino acid position numbers are shown.

FIG. 7 shows the sequence of the 5' end of the AmEPV NPH I gene including the deduced amino acids. The base numbers represent the extension of the sequence shown in FIG. 2, which includes the partial AmEPV NPH I (NTPase) gene. The sequence in FIG. 1 ends at base 6768. The base numbers correspond to those in FIG. 8.

FIG. 8 shows the spheroidin and NPH I gene orientations in AmEPV, and the locations of primers used in PCR reactions. The arrowheads over the numbers in the figure represent oligonucleotide primers. The arrowhead base shows the approximate starting point of the oligo and 5' to 3' is shown by the direction of the arrowhead.

RM53 and RM161 are AmEPV spheroidin specific primers. Primer RM212 is from the 3' end of the CbEPV NPH I gene and RM206 is just downstream of the CbEPV NPH I gene but within the published CbEPV sequence (Yuen et al. [1991], supra).

DETAILS OF DESCRIPTION OF THE INVENTION

The present invention provides novel Entomopoxvirus (EPV) polynucleotide sequences free from association with other viral sequences with which they naturally associated. Recombinant polynucleotide vectors comprising the sequences, recombinant viruses comprising the sequences, and host cells infected with the recombinant viruses are also disclosed herein. These compositions are useful in methods of the invention for the expression of heterologous genes and production of selected proteins in both insect and mammalian host cells.

Novel polynucleotide sequences of the invention encode EPV spheroidin genes and/or flanking sequences, including sequences which provide regulatory signals for the expression of the gene. The invention also provides novel polynucleotide sequences encoding an EPV thymidine kinase (tk) gene and/or its flanking sequences. The polynucleotide sequences of this invention may be either RNA or DNA sequences. More preferably, the polynucleotide sequences of this invention are DNA sequences.

Specifically disclosed by the present invention are spheroidin polynucleotide sequences obtainable from the Amsacta moorei Entomopoxovirus (AmEPV), Choristoneura biennis Entomopoxovirus (CbEPV), and Choristoneura fumiferana Entomopoxovirus (CfEPV). Also specifically exemplified is a tk polynucleotide sequence obtained from AmEPV. While these species are exemplified for practice of the methods and compositions of this invention, utilizing the techniques described herein, substantially homologous sequences may be obtained by one of ordinary skill in the art from other Entomopoxovirus species.

The AmEPV spheroidin DNA sequence, including flanking and regulatory sequences, is reported in FIG. 2 as spanning nucleotides #1 through 6768 (SEQ ID NO. 1). Within this sequence, the spheroidin gene coding sequence spans nucleotides #3080 through #6091 (SEQ ID NO. 21). A fragment which is likely to contain the promoter sequences spans nucleotides #2781 through #3199 (SEQ ID NO. 22). Other regions of that sequence have also been identified as putative coding regions for other structural or regulatory genes associated with spheroidin. These other fragments of interest include the following sequences: the
G2R ORF, which is nucleotides #1472 through #2151 (SEQ ID NO. 23) encoding the amino acid sequence shown in SEQ ID NO. 3; the G4 ORF, which is nucleotides #2302 through #2987 (SEQ ID NO. 24) encoding the amino acid sequence shown in SEQ ID NO. 5; and the following sequences transcribed left to right on FIG. 2: the G1L ORF, which is nucleotides #65 through #1459 (SEQ ID NO. 25) encoding the amino acid sequence shown in SEQ ID NO. 2; the G3L ORF, which is nucleotides #2239 through #2475 (SEQ ID NO. 26) encoding the amino acid sequence shown in SEQ ID NO. 4; and the G6 ORF, which includes nucleotides #6277 through #6768 (SEQ ID NO. 27) encoding the amino acid sequence shown in SEQ ID NO. 7. These ORFs are identified in FIG. 2. It should be noted that the full length of the G6 ORF extends beyond nucleotide #6768, is shown in SEQ ID NO. 1 and FIG. 7, and is discussed more fully below.

The AmEPV ORF G4R (SEQ ID NO. 24) which encodes G4R (SEQ ID NO. 5) is immediately upstream of the spheroidin gene has significant homology to the capripox

virus IM3 ORF. A homolog of the IM3 ORF is found in vacinia virus, just upstream of a truncated version of the cowpox virus ATl gene. Therefore, the microenvironments in this region are similar in the two viruses. Two other ORFs relate to counterparts in vaccinia virus. These ORFs include the 17 ORF of the vaccinia virus HindIII-I fragment (17)


As set out in detail in the accompanying examples below, the spheroidin gene of AmEPV was identified through direct microsequencing of the protein, and the results were used for the design of oligonucleotide probes. Transcription of the spheroidin gene is inhibited by cycloheximide, suggesting it is a late gene. Consistent with this prediction are the observations that spheroidin transcripts were initiated within a TATAAT motif (See FIG. 2, nucleotide #3077-3082) and that there is a 5' poly(A) sequence, both characteristic of late transcripts.

The isolation and sequencing of the ChEPV and CIEPV spheroidin genes are also described in detail below.

The AmEPV thymidine kinase (tk) DNA sequence, including flanking and regulatory sequences, is reported in FIG. 3, as spanning nucleotides #1 through #1511 (SEQ ID NO. 8). Within this sequence, the tk gene coding sequence spans nucleotides #234 through #782 (SEQ ID NO. 28) (transcribed right to left on FIG. 3). Another fragment of interest may include nucleotides #783 through #851 (SEQ ID NO. 29) of that sequence or fragments thereof. A fragment likely to contain the promoter regions spans nucleotides #750 through #890 (SEQ ID NO. 30). Other regions of that sequence have also been identified as putative coding regions for other structural or regulatory genes associated with tk. These other regions of interest include the following sequences (transcribed left to right on FIG. 3: ORF Q1, which is nucleotides #18 through #218 (SEQ ID NO. 31) encoding the amino acid sequence shown in SEQ ID NO. 10); and ORF Q3, which is nucleotides #852 through #1511 (SEQ ID NO. 32) encoding the amino acid sequence shown in SEQ ID NO. 10.

The location of the AmEPV tk gene maps in the EcoRI-Q fragment near the left end of the physical map of the AmEPV genome (FIG. 1) (see also, Hall, R. L., et al. [1990] Arch. Virol. 110:77-90, incorporated herein by reference). Because of the orientation of the gene within the AmEPV genome, transcription of the gene is likely to occur toward the terminus. There are believed to be similar tk genes, or variations thereof, in other systems, including mammalian systems. As set out in detail in the examples below, the tk gene of AmEPV was identified through direct microsequencing of the protein, and the results were used for the design of oligonucleotide probes.

The term “polynucleotide sequences” when used with reference to the invention can include the entire EPV spheroidin or tk genes with regulatory sequences flanking the coding sequences. The illustrated AmEPV sequences are also encompassed by that term. Also included in the definition are fragments of the coding sequences with flanking regulatory sequences. The definition also encompasses the regulatory sequences only, e.g., the promoter sequences, transcription sites, termination sequences, and other regulatory sequences.

Sequences of the invention may also include all or portions of the spheroidin or tk genes linked in frame to a heterologous gene sequence. Additionally, polynucleotide sequences of the invention may include sequences of the spheroidin or tk genes into which have been inserted a foreign or heterologous gene sequence, so that the EPV sequences flank the heterologous gene sequence.

Polynucleotide sequences of this invention also include sequences which are capable of hybridizing to the sequences of FIGS. 2 and 3, under stringent conditions. Also sequences capable of hybridizing to the sequences of FIGS. 2 and 3 under non-stringent conditions may fall within this definition providing that the biological or regulatory characteristics of the sequences of FIGS. 2 and 3, respectively, are retained. Examples of stringent and non-stringent conditions of hybridization are conventional (see, e.g., Sambrook et al. [1989] Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

Similarly, polynucleotide sequences of this invention also include variants, including allelic variations (naturally-occurring base changes in the EPV species population which may or may not result in an amino acid change) of DNA sequences encoding the spheroidin or tk protein sequences or other ORFs or regulatory sequences illustrated in FIGS. 2 and 3. Similarly, DNA sequences which encode spheroidin or tk proteins of the invention but which differ in codon sequence due to the degeneracies of the genetic code or variations in the DNA sequences which are caused by point mutations or by induced modifications to, for example, enhance a biological property or the usefulness of a desired polynucleotide sequence encoded thereby are also encompassed in the invention.

Utilizing the sequence data in FIGS. 2 or 3, as well as the denoted characteristics of spheroidin or thymidine kinase, it is within the skill of the art to obtain other DNA sequences encoding these polypeptides. For example, the structural gene may be manipulated by varying individual nucleotides, while retaining the same amino acid(s), or varying the nucleotides, so as to modify the amino acids, without loss of utility. Nucleotides may be substituted, inserted, or deleted.
by known techniques, including, for example, in vitro mutagenesis and primer repair.

The structural gene may be truncated at its 3'-terminus and/or its 5'-terminus. It may also be desirable to ligate a portion of the polypeptide sequence to a heterologous coding sequence, and thus to create a fusion peptide.

The polynucleotide sequences of the present invention may be prepared by a variety of techniques well known to those skilled in the art. The sequences may be prepared synthetically or can be derived from viral RNA or from available cDNA-containing plasmids by chemical and genetic engineering techniques or combinations thereof which are standard in the art.

The Entomopoxivirus proteins—spheroidin, thymidine kinase and their respective regulatory sequences, as described herein—may be encoded by polynucleotide sequences that differ in sequence from the sequences of Figs. 2 and 3 due to, for example, natural allelic or species variations. Thus, the terms spheroidin or tk polypeptides also refer to any of the naturally occurring sequences and variants thereof, e.g., processed or truncated sequences or fragments, including the mature spheroidin or tk polypeptides and mutant or modified polypeptides or fragments that retain the same utility and preferably have homology to Fig. 2 or 3, respectively, of at least 75%, more preferably 90%, and most preferably 95%.

Another aspect of the present invention is provided by the proteins encoded by the EPV spheroidin and tk polynucleotide sequences. Putative amino acid sequences of the two EPV proteins as well as additional putative proteins encoded by the ORFs of these sequences which are identified in Figs. 2 and 3, respectively. EPV spheroidin has no significant amino acid homology to any reportedly reported protein, including the polyhedrin protein of baculovirus. We have found that both spheroidin and tk are nonessential proteins, which makes them desirable as sites for insertion of exogenous DNA.

Comparison of the AmEPV tk amino acid sequence with other tk genes reveals that the AmEPV tk gene is not highly related to any of the vertebrate poxvirus tk genes (43.4-45.7%). The relatedness of the vertebrate tk proteins to AmEPV is still lower (39.3-41.0%), while African Swine Fever (ASF) showed the least homology of all the tk proteins tested (31.4%). Although ASF has many similarities to poxviruses, and both ASF and AmEPV infect invertebrate hosts, the tk genes indicate little commonality and/or indication of common origin stemming from invertebrate hosts.

The spheroidin and thymidine kinase polypeptide sequences may include isolated naturally-occurring spheroidin or tk amino acid sequences identified herein or deliberately modified sequences which maintain the biological or regulatory functions, or other utility, of the AmEPV polypeptides, respectively identified in Figs. 2 and 3. Therefore, provided that the utilities of these polypeptides are not destroyed despite such modifications, this invention encompasses the use of all amino acid sequences disclosed herein as well as variants thereof retaining spheroidin or tk utility. Similarly, proteins or functions encoded by the other spheroidin or tk ORFs may include sequences containing amino acid modifications but which retain their regulatory or other biological functions, or other utility.

Examples of such modifications include polypeptides with amino acid variations from the natural amino acid sequence of Entomopoxivirus spheroidin or thymidine kinase; in particular, conservative amino acid replacements. Conservative replacements are those that take place within a family of amino acids that are related in their side chains.

Genetically encoded amino acids are generally divided into four families: (1) acidic=aspartate, glutamate; (2) basic=lysine, arginine, histidine; (3) non-polar=alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar=glutamine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanne, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an asparagine with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid will not have a major effect on biological activity, especially if the replacement does not involve an amino acid at an active site of the polypeptides.

As used herein, the term “polypeptide” refers to a polymer of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. The phrase “polypeptide and variants thereof” includes post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The proteins or polypeptides of the present invention may be expressed in host cells and purified from the cells or media by conventional means (Sambrook et al., supra).

This invention also relates to novel viral recombinant polynucleotide molecules or vectors, which permit the expression of heterologous genes in a selected host cell. Such a polynucleotide vector of the invention comprises the polynucleotide sequence encoding all or a portion of the spheroidin or tk gene, the RNA polymerase from a selected poxvirus, and the polynucleotide sequence encoding a desired heterologous gene. Preferably, the sequence includes the regulatory region, and most preferably, the promoter region, of either the EPV spheroidin or tk gene. In addition, the source of the polymerase is not limited to EPV; rather, any poxvirus RNA polymerase may be utilized.

Therefore, the viral vectors may contain other viral elements contributed by another poxvirus, either vertebrate or invertebrate, with the only EPV sequences being provided by the presence of the EPV spheroidin or tk gene sequences, or fragments thereof. Numerous conventional expression viral vectors and expression systems are known in the art. Particularly desirable vectors systems include those of vertebrate or invertebrate poxviruses. The Entomopoxivirus spheroidin and tk gene regulatory sequences may be used in other virus vector systems which contain a poxvirus RNA polymerase to enhance the performance of those systems, e.g., in vaccinia vectors. Methods for the construction of expression systems, in general, and the components thereof, including expression vectors and transformed host cells, are within the art. See, generally, methods described in standard texts, such as Sambrook et al., supra. The present invention is therefore not limited to any particular viral expression system or vector into which a polynucleotide sequence of this invention may be inserted, provided that the vector or system contains a poxvirus RNA polymerase.

The vectors of the invention provide a helper independent vector system, that is, the presence or absence of a functional spheroidin or tk gene in a poxvirus contributing elements to the vector, e.g., contributing the RNA polymerase, does not affect the usefulness of the resulting recombinant viral vector. Because both spheroidin and tk are non-essential
genes, the viral vectors of this invention do not require the presence of any other viral proteins, which in helper-dependent systems are often contributed by additional viruses to coinherit the selected host cell.

Selected host cells which, upon infection by the viral vectors will permit expression of the heterologous gene, include insect and mammalian cells. Specifically, if the viral vector comprises the EPV spheroidin or tk gene sequences of the invention inserted into any member of the family Entomopoxvirinae, e.g., EPVs of any species, the host cell will be limited to cells of insects normally infected by EPVs. If the viral vector comprises the EPV spheroidin or tk gene sequences of the invention inserted into a vertebrate poxvirus, such as vaccinia or swinepox, the host cells may be selected from among the mammalian species normally infected with the wild-type vertebrate poxvirus. Most desirably, such mammalian cells may include human cells, rodent cells and primate cells, all known and available to one of skill in the art.

According to one aspect of the subject invention, therefore, vectors of the present invention may utilize a fragment of the polynucleotide sequence of EPV spheroidin, particularly the promoter and ancillary regulatory sequences which are responsible for the naturally high levels of expression of the gene. Desirably, spheroidin sequences may be found within the sequence of FIG. 2 (SEQ ID NO. 1), more particularly within the region of nucleotides 2781 through 3199 (SEQ ID NO. 22). Smaller fragments within that region may also provide useful regulatory sequences. The desired spheroidin promoter sequence can be utilized to produce large quantities of a desired protein by placing it in an operative association with a selected heterologous gene in viral expression vectors capable of functioning in either a vertebrate or invertebrate host cell.

As used herein, the term "operative association" defines the relationship between a regulatory sequence and a selected protein gene, such that the regulatory sequence is capable of directing the expression of the protein in the appropriate host cell. One of skill in the art is capable of operatively associating such sequences by resort to conventional techniques,

Where the spheroidin polynucleotide sequence in the vector contains all or a portion of the spheroidin coding sequence in association with, or linked to, the heterologous gene sequences, the resulting protein expressed in the host cell may be a fusion protein consisting of all or a portion of the spheroidin protein and the heterologous protein. Where the spheroidin polynucleotide sequence in the vector does not contain sufficient coding sequence for the expression of a spheroidin protein or peptide fragment, the heterologous protein may be produced alone.

In an analogous manner, the promoter and regulatory sequences of tk (FIG. 3 SEQ ID NO. 8) may be employed in the construction of an expression vector to drive expression of a heterologous protein, or a fusion protein, in a selected known expression system. These tk regulatory sequences are desirably obtained from the sequence of FIG. 3 (SEQ ID NO. 8), particularly in the fragment occurring between nucleotide 750 through 890 (SEQ ID NO. 30). Smaller fragments within that region may also provide useful regulatory sequences.

An advantage of the use of the novel EPV spheroidin or tk promoter sequences of this invention is that these regulatory sequences are capable of operating in a vertebrate poxvirus (e.g., vaccinia)-mammalian cell expression vector system. For example, the strong spheroidin promoter can be incorporated into the vaccinia virus system through homologous recombination. Unlike the promoter for the baculovirus polyhedrin gene, the promoter for the EPV spheroidin gene can be utilized directly in the vaccinia or swinepox virus expression vector.

To construct a vector according to the present invention, the spheroidin or tk polynucleotide sequence may be isolated and purified from a selected Entomopoxvirus, e.g., AmEPV, and digested with appropriate restriction endonuclease enzymes to produce a fragment comprising all or part of the spheroidin or tk gene. Alternatively such a fragment may be chemically synthesized.

Still alternatively, the desired AmEPV sequences may be obtained from bacterial cultures containing the plasmids pH512, pMEG3k-1 or pRH7. The construction of the plasmid pH512 is described in the examples below. This plasmid contains the 4.51 kb BglII fragment AmEPV DNA sequence inserted into a BamHI site in the conventional vector pUC9. The plasmid pRH7 was constructed by digesting AmEPV genomic DNA, obtained as described in Example 1, with Bgl2I, and the resulting fragments with HaeIll. T4 DNA polymerase is employed to blunt end the AmEPV DNA and the fragment containing the spheroidin gene is ligated to the large fragment of a Smal digested pUC9 fragment. This fragment contains the entire spheroidin open reading frame and some flanking sequence, included within the nucleotide sequence spanning #2274-6182 (SEQ ID NO. 33) of FIG. 2. The construction of plasmid pMEG3k-1 comprising the regulatory sequences of the tk gene as well as the structural gene is described below in the Example 8. It was constructed by inserting the EcoRI-Q fragment of AmEPV into the conventional vector pUC18.

Bacterial cultures containing plasmids pH512, pMEG tk-1, and pRH7 have been deposited in the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Md., USA. The deposited cultures are as follows:

<table>
<thead>
<tr>
<th>Culture</th>
<th>Accession No.</th>
<th>Deposit Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli SURE strain (Strainage) pMEG-kl</td>
<td>ATCC 68532</td>
<td>February 26, 1991</td>
</tr>
<tr>
<td>E. coli SURE strain (Strainage) pH512</td>
<td>ATCC 68533</td>
<td>February 26, 1991</td>
</tr>
<tr>
<td>E. coli SURE strain (Strainage) pRH7</td>
<td>ATCC 68902</td>
<td>January 28, 1992</td>
</tr>
</tbody>
</table>

The plasmids can be obtained from the deposited bacterial cultures by use of standard procedures, for example, using cleared lysate-isopycnic density gradient procedures, and the like.

These ATCC deposits were made under conditions that assure that access to the cultures will be available during the pendency of this patent application to one determined by the Commissioner of Patents and Trademark to be entitled thereto under 37 CFR 1.14 and 35 USC 122. The deposits will be available as required by foreign patent laws in countries wherein counterparts of the subject application, or its progeny, are filed. However, it should be understood that the availability of a deposit does not constitute a license to practice the subject invention in derogation of patent rights granted by government action.

Further, the subject culture deposits will be stored and made available to the public in accord with the provisions of the Budapest Treaty for the Deposit of Microorganisms, i.e., they will be stored with all the care necessary to keep them viable and uncontaminated for a period of at least five years after the most recent request for the furnishing of a sample.
of a deposit, and in any case, for a period of at least 30 (thirty) years after the date of deposit or for the enforceable life of any patent which may issue disclosing the cultures. The depositor acknowledges the duty to replace the deposit(s) should the depository be unable to furnish a sample when requested, due to the condition of a deposit. All restrictions on the availability to the public of the subject culture deposits will be irrevocably removed upon the granting of a patent disclosing them.

The molecular biology procedures referred to herein in describing construction of the vectors of this invention are standard, well-known procedures. The various methods employed in the preparation of the plasmid vectors and transformation or infection of host organisms are well-known in the art. These procedures are all described in, for example, Sambrook et al., cited above. Thus, it is within the skill of those in the genetic engineering art to extract DNA from microbial cells, perform restriction enzyme digestions, electrophorese DNA fragments, tail and anneal plasmid and insert DNA, ligate DNA, transform cells, prepare plasmid DNA, electrophorese proteins, and sequence DNA.

Because the AmEPV genome has no known unique restriction sites into which selected genes may be effectively introduced in a site-specific manner so as to be under the control of the spheroiodin or tk promoter sequences, such restriction sites must be introduced into desired sites in the selected EPV polynucleotide sequence. For example, the unique BstBI site located at nucleotide #3172 downstream of the start of the spheroiodin gene is the closest site to genetically engineer a usable insertion sequence for cloning. Therefore, restriction sites closer to the initiating Met of the spheroiodin gene must be deliberately inserted.

Methods for the insertion of restriction sites are known to those of skill in the art and include, the use of an intermediate shuttle vector, e.g., by cloning the EPV sequence into the site of an appropriate cloning vehicle. It will be recognized by those skilled in the art that any suitable cloning vehicle may be utilized provided that the spheroiodin or tk gene and flanking viral DNA may be functionally incorporated.

A spheroiodin shuttle vector may be constructed to include elements of the spheroiodin structural gene, a cloning site located or introduced in the gene to enable the selected heterologous gene to be properly inserted into the viral genome adjacent to, and under the control of, the spheroiodin promoter, and flanking viral DNA linked to either side of the spheroiodin gene to facilitate insertion of the spheroiodin-foreign gene-flanking sequence into another expression vector. The presence of flanking viral DNA also facilitates recombination with the wild type Enomopoxivirus, allowing the transfer of a selected gene into a replicating viral genome.

The shuttle vectors may thereafter be modified for insertion of a selected gene by deleting some or all of the sequences encoding for spheroiodin or tk synthesis near the respective transcriptional start sites. Examples of such sites in spheroiodin are nucleotides #3077 and 3080 and in tk includes nucleotide #809. Conventional procedures are available to delete spheroiodin or tk coding sequences.

As an alternative to or in addition to the restriction site, a variety of synthetic or natural oligonucleotide linker sequences may be inserted at the site of the deletion. A polynucleotide linker sequence, which may be either a natural or synthetic oligonucleotide, may be inserted at the site of the deletion to allow the coupling of DNA segments at that site. One such linker sequence may provide an appropriate space between the two linked sequences, e.g., between the promoter sequence and the gene to be expressed. Alternatively, this linker sequence may encode, if desired, a polypeptide which is selectively cleavable or digestible by conventional chemical or enzymatic methods. For example, the selected cleavage site may be an enzymatic cleavage site, including sites for cleavage by a proteolytic enzyme, such as enterokinase, factor Xa, trypsin, collagenease and thrombin. Alternatively, the cleavage site in the linker may be a site capable of being cleaved upon exposure to a selected chemical, e.g., cyanogen bromide or hydroxylamine. The cleavage site, if inserted into a linker useful in the sequences of this invention, does not limit this invention. Any desired cleavage site, of which many are known in the art, may be used for this purpose. In another alternative, the linker sequence may encode one or a series of restriction sites.

It will be recognized by those skilled in the art who have the benefit of this disclosure that linker sequences bearing an appropriate restriction site need not be inserted in place of all or a portion of the spheroiodin structural sequence, and that it would be possible to insert a linker in locations in the Enomopoxivirus genome such that both the sequence coding for the selected polypeptide and the spheroiodin structural sequence would be expressed. For instance, the sequence coding for the selected polypeptide could be inserted into the tk gene in place of all or a portion of the tk structural sequence and under the transcriptional control of the tk promoter.

Polymerase chain reaction (PCR) techniques can also be used to introduce convenient restriction sites into the EPV DNA, as well as to amplify specific regions of the EPV DNA. These techniques are well known to those skilled in this art. See, for example, PCR Protocols: A Guide to Methods and Applications, M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, (1990).

By use of these techniques, a variety of alternative modified shuttle vectors into which a selected gene or portion thereof may be incorporated may be suitably utilized in the present invention.

As one embodiment of the invention, therefore, the polynucleotide sequence, described above, may be used as a shuttle vector to transfer a selected heterologous gene to a selected virus. In this embodiment, the polynucleotide sequence encoding the EPV spheroiodin gene or EPV tk gene, or a fragment thereof, is linked to a heterologous gene. The polynucleotide sequence further contains a flanking region on either side of the spheroiodin-heterologous gene or tk-heterologous gene to enable ready transfer into a selected virus. This resulting construct is termed a cassette. Such a flanking region may be derived from EPV, or alternatively, may be complementary to the target virus. For example, if it is desirable to insert a selected heterologous gene into a vaccinia virus to create a recombinant virus, one would utilize flanking regions complementary to the targeted vaccinia virus. Similarly if the heterologous gene is inserted within the EPV spheroiodin or tk gene, so that the selected EPV regulatory sequence and heterologous gene are flanked by the EPV gene's own sequences, this cassette may be used for transfer into a wild type EPV having homologous sequences to the flanking sequences.

The insertion or linkage of the foreign gene into the tk or spheroiodin sequences of the present invention or the linkage of flanking sequences foreign to the spheroiodin or tk genes may be accomplished as described above. The vectors of the subject invention may use cDNA clones of foreign genes, because poxvirus genes contain no introns, presumably as a
consequence of a totally cytoplasmic site of infection.

In accordance with standard cloning techniques, any selected gene may be inserted into the vector at an available restriction site to produce a recombinant shuttle vector. Virtually any gene of interest could be inserted into the vectors described herein in order to obtain high expression of the desired protein. The, spero-90idin gene product may be useful as a particulate biological carrier for foreign gene antigens. Thus, a foreign gene fused to the spero-91idin gene may be useful as a method to produce a foreign protein attached to an effective vaccine carrier. Restriction sites in the fragment may thereafter be removed so as to produce a preferred spero-92idin or tk shuttle vector, having one or more cleavage or cloning sites located in the 3' direction downstream from the spero-93idin promoter sequence. Thus, the present invention is not limited by the selection of the heterologous gene.

Alternatively, a vector of this invention may comprise a heterologous gene which is inserted into all or a portion of the EPV spero-94idin or tk protein encoding sequence to interrupt the protein's natural processing. However, when the vector is transferred to another virus which contains a wild-type spero-95idin or tk gene, expression of the inserted heterologous gene is obtained. Thus, the Entomopoxivirus spero-96idin gene (FIG. 2; SEQ ID NO. 1) and/or the tk gene (FIG. 3; SEQ ID NO. 8) can be used as the location for the insertion of exogenous (heterologous) DNA in any of the above-mentioned expression systems. A shuttle vector so constructed may be useful as a marker for research and production techniques for identifying the presence of successfully integrated heterologous genes into the selected expression system.

The tk gene is a particularly desirable site for insertion of a selected heterologous gene. Unlike spero-97idin, tk is produced early in infection and in lesser quantities. Additionally, many poxviruses possess tk genes which may be sufficiently homologous to the EPV tk to provide easy recombination. For example, in vaccinia virus expression systems for mammalian cells, the vaccinia tk gene is a common insertion site. Therefore, the use of this gene is particularly desirable for construction of a shuttle vector to shuttle selected genes directly between vector systems. More specifically, a foreign gene may be desirably inserted into the EPV tk gene sequence between nucleotide #460 and #560 (See FIG. 3).

Insertion of cassettes containing foreign genes into wild-type poxviruses can be accomplished by homologous recombination. The homologous recombination techniques used to insert the genes of interest into the viruses of the invention are well known to those skilled in the art. The shuttle vectors, when co-infected into host cells with a wild-type virus, transfer the cassette containing the selected gene into the virus by homologous recombination, thereby creating recombinant virus vectors.

Expression of a selected gene is accomplished by infecting susceptible host insect cells with the recombinant viral vector of this invention in an appropriate medium for growth. An EPV expression vector is propagated in insect cells or insects through replication and assembly of infectious virus particles. These infectious vectors can be used to produce the selected gene in suitable insect cells, thus facilitating the efficient expression of the selected DNA sequence in the infected cell. If the EPV spero-98idin gene (or tk gene)-heterologous gene fragment is inserted into a vertebrate poxvirus by the same methods as described above, the recombinant virus may be used to infect mam-
The preferred recombinant AmEPV expression vector, comprising a hybrid spheroidin-IFN-β gene incorporated into the AmEPV genome, can thereafter be selected from the mixture of nonrecombinant and recombinant Entomopoxviruses. The preferred, but by no means only, method of selection is by screening for plaques formed by host insect cells infected with viruses that do not produce viral occlusions. Selection may be performed in this manner because recombinant EPV viruses which contain the spheroidin protein coding sequences interrupted by the heterologous gene are defective in the production of viral occlusions due to the insertional inactivation of the spheroidin gene.

Also, the selection procedure may involve the use of the β-galactosidase gene to facilitate color selection. This procedure involves the incorporation of the E. coli β-galactosidase gene into the shuttle vector and is well known to those skilled in the art. This technique may be of particular value if the exogenous DNA is inserted into the tk gene so that the spheroidin gene is still expressed. It will be recognized by those skilled in the art that alternative selection procedures may also be utilized in accordance with the present invention.

Accordingly, the DNA from a recombinant virus is thereafter purified and may be analyzed with appropriate restriction enzymes, for example, or PCR technology, to confirm that the recombinant AmEPV vector has an insertion of the selected gene in the proper location.

The vectors and methods provided by the present invention are characterized by several advantages over known vectors and vector systems. Advantageously, such EPV viral vectors of the present invention are not oncogenic or tumorigenic in mammals. Also, the regulatory signals governing Amosca moorei Entomopoxvirus (AmEPV) gene expressions are similar to those of vaccinia. Therefore, it is possible to transfer the strongly expressed spheroidin gene, or the thymidine kinase gene, as an expression cassette, not only in insect cells, but for use in vertebrate poxviruses such as vaccinia and swinepox virus.

Based on reported data with vaccinia, herpes and baculovirus vector systems, which suggest that up to 30 kb can be transferred without disrupting the vector viability, the normal limitation on the amount of exogenous DNA which can be packaged into a virus is not anticipated to be encountered when using the novel EPV vectors and methods of the subject invention.

Another advantage is that for the novel vectors of the subject invention, the transcription and translation of foreign proteins is totally cytoplasmic. Still another advantage lies in the expression power of the EPV spheroidin regulatory sequences, which when in operative association with a heterologous gene in a vector of this invention, can be used to produce high levels of heterologous protein expression in the selected host cell.

The EPV vectors of this invention and methods for employing them to express selected heterologous proteins in insect or mammalian cells, as described above, are characterized by the advantage of replication in insect cells, which avoids the use of mammalian viruses, thereby decreasing the possibility of contamination of the product with mammalian virus. The expression system of this invention is also a helper independent virus expression vector system. These two characteristics are shared by known baculovirus expression systems. However, as shown in Table 1, the EPV expression vector system (EEVS) using the vectors of this invention has some important distinguishing features compared to the baculovirus expression systems (BEVS).

### Table 1: Differences between EEVS and BEVS

<table>
<thead>
<tr>
<th>Site of replication:</th>
<th>cytoplasm</th>
<th>nucleus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virus family:</td>
<td>Poxviridae</td>
<td>Baculoviridae</td>
</tr>
<tr>
<td>Sites for insertion of foreign genes:</td>
<td>spheroidin &amp; thymidine kinase</td>
<td>polyhedrin &amp; p10</td>
</tr>
<tr>
<td>Shuttle possibilities between vertebrate and insect systems:</td>
<td>Orthopoxviruses</td>
<td>No mammalian counterparts.</td>
</tr>
<tr>
<td></td>
<td>Leporipoxviruses</td>
<td>Baculovirus is not known to contain a tk gene.</td>
</tr>
<tr>
<td></td>
<td>Suipoxviruses</td>
<td>Polyhedrin is not found in mammalian systems.</td>
</tr>
</tbody>
</table>

The present invention also provides a method for screening recombinant host cells for insertion of heterologous genes by use of the recombinant viral polynucleotide molecules of this invention. The viral molecules containing the selected heterologous gene sequence linked to the polynucleotide sequence encoding less than all of the Entomopoxvirus spheroidin protein. The heterologous gene may be linked to the spheroidin or tk regulatory sequences in the absence of the complete coding sequences, or it may be inserted into the spheroidin or tk gene coding sequences, thus disrupting the coding sequence. The cell infected with the recombinant vector is cultured under conditions suitable for expression of the heterologous protein, either fused or as a fusion protein with a portion of the spheroidin sequence.

The absence of occlusion bodies which would ordinarily be formed by the expression of the intact spheroidin protein indicates the integration of the heterologous gene.

If the viral vector similarly contained either incomplete or interrupted EPV tk encoding sequence, the absence of thymidine kinase function (e.g., resistance to methotrexate or an analogue thereof) formed by the integration of the inactive thymidine kinase sequence indicates the insertion of the heterologous gene.

Alternatively, if a parent virus is deleted of part of its tk or spheroidin gene, and is thereafter mixed with a viral vector containing intact tk or spheroidin fused to the foreign gene, recombinants would express the methotrexate resistance or produce occlusion bodies, respectively, thus indicating integration of the active tk or spheroidin genes and the foreign gene.

The above-described selection procedures provide effective and convenient means for selection of recombinant Entomopoxvirus expression vectors.

Another embodiment of the present invention involves using novel EPV expression systems of the subject invention for insect control. Control of insect pests can be accomplished by employing the vectors and methods of the invention as described above. For example, a gene coding for an selected insect toxin may be inserted into the viral expression vector under the control of the spheroidin or tk regulatory sequences or within either of the two genes for purposes of recombination into a selected virus having homologous flanking regions.

Genes which code for insect toxins are well known to those skilled in the art. An exemplary toxin gene isolated from Bacillus thuringiensis (B.t.) can be used according to the subject invention. B.t. genes are described, for example, in U.S. Pat. Nos. 4,775,131 and 4,865,981. Other known insect toxins may also be employed in this method.
The resulting EPV vector containing the toxin gene is applied to the target pest or its surroundings. Advantagesously, the viral vector will infect the target pest, and large quantities of the toxin will be produced, thus resulting in the control of the pest. Particularly large quantities of the toxin protein can be produced if the regulatory sequences of the Entomopoxovirus spheroidin gene are used to express the toxin.

Alternatively, the spheroidin gene can be left intact and the toxin gene inserted into a different Entomopoxovirus gene such as the tk gene. In this construct, the toxin will be produced by the system and then effectively coated or encapsulated by the natural viral production of spheroidin. This system thus produces a toxin which will advantageously persist in the environment to prolong the availability to the target pest.

In addition to the novel Entomopoxovirus expression vectors and methods for their use described herein, the subject invention pertains to the use of novel regulatory elements from Entomopoxovirus to construct novel chimeric vaccinia and swinepox vaccines and expression systems which are functional across genera of mammalian poxviruses. The polynucleotide sequences of the invention can also be used with viral vaccines, e.g., known vaccinia virus vaccines, to enhance the effectiveness of these vaccines. Such vaccines have been described for use in controlling rabies and other infectious diseases in mammals. Specifically, it is anticipated that the introduction of the EPV spheroidin promoter sequences into known viral vectors which are used to express selected proteins in a mammalian host in vivo may enable the powerful spheroidin promoter to increase expression of the protein in the viral vaccine. This aspect of the invention provides a significant improvement over other expression systems, including the baculovirus expression system (BEVS).

The following examples illustrate the compositions and procedures, including the best mode, for practicing the invention. These examples, should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted. The restriction enzymes disclosed herein can be purchased from Bethesda Research Laboratories, Gaithersburg, Md., or New England Biolabs, Beverly, Mass. The enzymes are used according to the instructions provided by the supplier. Klenow fragment of DNA polymerase, T4 polynucleotide kinase, and T4 DNA ligase were obtained from New England Biolabs and Promega.

Example 1
Preparation of AmEPV DNA

The replication of AmEPV has been described previously (Goodwin, R. H., et al. [1990] J. Invertebr. Pathol, 56:190–205). The gypsy moth (Lymantria dispar) cell line IPLB-LD-652 (Insect Pathology Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Md.) is maintained at 26° to 28°C in EX-CELL 400 (JKH Biosciences, Lenexa, Kans.) supplemented with 10% fetal bovine serum, 100 U of penicillin, and 100 μg of streptomycin per ml. Other insect cell lines are well known to those skilled in the art and can be used according to the subject invention.

The AmEPV inoculum for cell culturing was from an AmEPV-infected, freeze-dried E. acrae larva stored at ~70°C. (Hall, R. L., et al. [1990] Arch. Virol. 110:77–90). The larva was crushed and macerated in 5 ml of EX-CELL 400 (with penicillin and streptomycin but without fetal bovine serum) to which 0.003 g of cysteine-HCl had been added to prevent melanization. The debris was pelleted at 200g for 5 minutes, and the supernatant was passed through a 0.45-μm-pore-size filter.

The gypsy moth cells were infected with AmEPV by addition of the inoculum to a preconfluent monolayer of cells (about 0.1 to 1 PFU per cell), with occasional agitation of the dish during the first day. Infected cells were harvested 5 to 6 days postinfection.

AmEPV DNA was prepared from the infected cells by one of two methods. The first method involved in situ digestion of infected cells embedded within agarose plugs, after which the released cellular and viral DNAs were separated by pulsed-field electrophoresis (Bio-Rad CHEF-2 DR-system). IPLB-LD-652 cells were infected with first-cell-culture-passage AmEPV. Infected cells were harvested 6 days postinfection by centrifugation at 2000g for 5 minutes, rinsed, and resuspended in modified Hank's phosphate-buffered saline (PBS), which contained 15 g of glucose per liter, but no Ca2+ or Mg2+.

For embedding of the infected cells in agarose plugs, 1% SeaPlaque GTG agarose (prepared in modified Hank's PBS and equilibrated at 37°C) was mixed 1:1 with infected cell to yield 5×10⁶ cells per ml in 0.5% agarose. Digestion to release DNA was done by gentle shaking of the inserts in 1% Sarkosyl-0.5M EDTA-1 mg of proteinase K per ml at 50°C for 2 days (Smith, C. L., et al. [1987] Methods Enzymol. 151:461–489). The CHEF-2 DR parameters for DNA separation were 180 V, a pulse ratio of 1, 50 initial and 90 second final pulse times, and a run time of 20 to 25 hours at 4°C. The separating gel was 1% SeaKem GTG agarose in 0.5x TBE buffer (Sambrook et al., supra). Viral DNA bands were visualized by ethidium bromide staining and electroeluted (Allington, W. B., et al. [1978] Anal. Biochem. 85:188–196). The recovered DNA was used for plasmid cloning following ethanol precipitation.

The second method of viral DNA preparation used the extracellular virus found in the infected-cell-culture supernatant. The supernatant from 10-day-postinfection cell cultures was clarified by centrifugation at 200g for 5 min. Virus was collected from the supernatant by centrifugation at 12,000g. Viral pellets were resuspended in 6 ml of 1x TE. DNase I and RNase A (10 and 20 μg/ml final concentrations, respectively) were added, and the mixture was incubated at 37°C for 30 minutes. The mixture was heated to 50°C for 15 minutes. SDS and proteinase K (1% and 200 μg/ml, respectively) were then added. The sample was incubated to 50°C overnight and extracted three times with buffer-saturated phenol and once with SEVAG (Sambrook et al., supra). The DNA was ethanol precipitated and resuspended in 1x TE (pH 8).

For routine virus quantitation, 1 ml of an appropriate virus dilution (prepared in un-supplemented EX-CELL 400) was added to a preconfluent monolayer of cells in a 60 mm culture dish, with intermittent agitation over a 5 hour adsorption period at 26° to 28°C. The virus inoculum was removed, and 5 ml of a 0.75% SeaPlaque agarose (PMC BioProducts, Rockland, Me.) overlay prepared with 2x EX-CELL 400 and equilibrated at 37°C was added to the monolayer. Plaques were visualized after 5 days of incubation at 26°C by inspection with a stereomicroscope.

The DNA prepared according to either method was then cut with a variety of restriction endonuclease enzymes, e.g., BamHI, EcoRI, HindIII, PstI and Xbol, generating the various fragments which appear on the physical map of FIG. 1. Hereafter, reference to each restriction fragment will refer to the enzyme and the applicable letter, e.g., BamHI-A.
through BamHI-E, EcoRI-A through EcoRI-S, etc.

Production of Spheroidin Polypeptide

To localize the spheroidin gene, a purified preparation of occlusion bodies (OBs) from infected caterpillars was solubilized and subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, J. K., 1970) *Nature* (London) 227:680-685) with a 4% acrylamide stacking gel and a 7.5% separating gel. The acrylamide used to prepare spheroidin for protein microsequencing was deionized with AG501X8 resin (Bio-Rad, Richmond, Calif.). The gels were polymerized overnight at 4°C. For sample preparation, 2x Laemmli sample buffer consisting of 125 mM Tris-HCl (pH 6.8), 4% SDS (w/v), 10% β-mercaptoethanol (v/v), and 20% glycerol (v/v) was used.

OB suspension samples were diluted 1:1 with 2x Laemmli sample buffer and boiled for 5 minutes. Several lanes of an OB protein preparation were separated electrophoretically. The spheroidin protein (113 kDa) was the predominant protein of the purified OBs. Spheroidin within SDS-polyacrylamide gels was tested for glycosylation by periodic acid-Schiff staining (Zacharius, R. M., et al. [1969] *Anal. Biochem.* 30:149–152).

Following electrophoretic separation, several lanes in the unstained gel were transferred to an Immobilon polyvinylidene difluoride (PVDF) membrane with a Bio-Rad TransBlot apparatus at 90 V for 2 hours in a buffer consisting of 10 mM morpholinepropanesulfonic acid (pH 6.0) and 20% methanol. Spheroidin was visualized on the PVDF membrane by Coomassie blue staining.

The region of the PVDF membrane containing spheroidin was excised from the membrane, and direct protein microsequencing was done with an Applied Biosystems gas-phase sequencer. Microsequencing of the intact protein was unsuccessful, presumably because the N terminus of the protein was blocked.

Cyanogen bromide cleavage was performed on samples of spheroidin eluted from the PVDF membrane to generate internal peptide fragments for sequencing. Major polypeptides of 15, 9, 8, and 6.2 kDa were produced.

Example 3

Sequencing, Hybridizations

All DNA sequencing was done by the dideoxy chain termination method (Sanger, F., et al. [1977] *Proc. Natl. Acad. Sci. USA* 74:5463–5467) with [α-35S]dATP and Sequenase (US Biochemical, Cleveland, Ohio). Standard sequencing reactions with Sequenase were carried out in accordance with the instructions of the manufacturer, US Biochemical.

A reliable amino acid sequence was obtained from the 9, 8, and 6.2 kDa polypeptides produced as described in Example 2. The 8 and 9 kDa polypeptides represented overlapping partial CNBr cleavage products which together yielded the longest continuous amino acid sequence: Met-Ala-(Asn or Arg)-Asp-Leu-Val-Ser-Leu-Leu-Phe-Met-(Asn or Arg)-(?)-Tyr-Val-(Asn?)-(Ile-Glu-Ile-Asn-Glu-Ala-Val-?) (SEQ ID NO. 34). The amino acid sequence obtained from the 6.2 kDa fragment was Met-Lys-Ile-Thr-Ser-Thr-Glu-Val-Asp-pro-Glu-Tyr-Val-(Thr-Prole)-Ser-(Asn?) (SEQ ID NO. 35). A partial sequence for the 15 kDa fragment was also obtained: (Asn?)-Ala-Leu-Phe-(Phe?)-(Asn?)-(Val-Phe) (SEQ ID NO. 36). The question marks in the above sequences indicated undetermined or unconfirmed amino acids. All sequences were ultimately located within the spheroidin gene sequence.

Example 4

Plasmid pH512

A BglII AmEPV DNA library was prepared by digesting the genomic AmEPV DNA with BglII according to manufacturer’s instructions. Plasmid pUC9 ( Gibco ; Bethesda Research Labs) was BglII-digested and phosphatase-treated. The genomic BglII cut AmEPV was shotgun cloned into the BamHI site of pUC9. *Escherichia coli* strain SURE (Stratagene, La Jolla, Calif.) was transformed by electroporation with a Bio-Rad Gene Pulser following the instructions provided by the manufacturer with the shotgun ligation, containing a variety of recombinant plasmids. Mini-preparations of plasmids were made by a conventional alkaline lysis procedure (Sambrook et al., supra). These plasmids were cut with EcoRI-SalI to release the insert and run on a gel. The resulting plasmid DNA was southern blotted to a nylon membrane, producing a number of clones.

Among the fragments produced from the restriction enzyme digests of the genomic DNA was a 4.4 BglII fragment and an EcoRI-D fragment. In order to locate a desirable clone from among those produced above, the sequence derived from the 6.2 kDa CNBr fragment was used to design a degenerate oligonucleotide for use as a hybridization probe to locate the spheroidin gene in a clone. The nucleotide sequence of this probe called RM58 (SEQ ID NO. 12) was GAAATGCATCCTGGAATAGT (5' to 3') which represents A or G, 6 represents C or T, and 7 represents A, G, C, or T. The peptide sequence of the probe was: Glu-Val-Asp-Pro-Glu-Tyr (SEQ ID NO. 37).

The DNA probe was radiolabeled either with [α-32P]dCTP by the random oligonucleotide extension method (Feinberg, A. P., et al. [1983] *Anal. Biochem.* 132:6–13) or with [γ-32P]ATP and T4 polynucleotide kinase (Sambrook et al., supra). These same procedures were used for all other oligonucleotide probes described below. Both types of probes were purified by passage through spin columns of Sephadex G-50.

Southern transfer was done with Hybond-N (Amersham); the transferred DNA was fixed to the membrane by UV cross-linking. Southern hybridization was performed both with transfected DNA including the restriction fragments described above, as well as the BglII library of AmEPV DNA cloned into BamHI-digested plasmid pUC9 as described above. Hybridization with the oligonucleotide probe was done at 37°C or 45°C with BLOTTO (Sambrook et al., supra) and was followed by two washes at room temperature with 0.3M NaCl-0.06M Tris (pH 8)-2 mM EDTA for 5 minutes.

The RM58 probe (SEQ ID NO. 12) hybridized to the 4.4 kb BglII fragment and the EcoRI-D fragment of AmEPV DNA (See FIG. 1). A plasmid produced by the shotgun cloning, recombinant pH512 (a BglII 4.56 kb fragment inserted into the BamHI site of pUC9 which contains about 1.5 kb of the 5' end of the spheroidin gene) was also identified by this hybridization with the RM58 oligonucleotide (SEQ ID NO. 12).

The 4.51 kb pH512 BglII insert was isolated, radiolabeled as described above, and hybridized back to various AmEPV genomic digests as follows. The DNA—DNA hybridization was done at 65°C with BLOTTO (Sambrook et al., supra) and was followed by two washes at room temperature with 0.3M NaCl-0.06M Tris (pH 8)-2 mM
EDTA for 5 minutes, two washes for 15 minutes each at 65°C. but with 0.4% SDS added, and two washes at room temperature with 0.03M NaCl-0.06M Tris (pH 8)-0.2 mM EDTA. Hybridization was observed to the BamHI-A,
EcoRI-D, HindIII-G and J, PsiI-A, and Xhol-B fragments of AmEPV DNA. The results of these hybridizations indicated that the 4.51 kb fragment in pRH512 was substantially identical to the 4.4 kb fragment produced by BglII digestion of genomic DNA.

The 4.51 kb BglII insert of pRH512 was thereafter sequenced by two procedures. One is the double-stranded plasmid sequencing method (Hattori, M., et al. [1986]Anal. Biochem. 152:232–238) performed with "miniprep" (Sambrook et al., supra) DNA and 1 pmol of universal, reverse, or custom-designed oligonucleotide primer in each sequencing reaction. Nested exonuclease II deletions (Henikoff, S. [1987] Methods Enzymol. 155:156–165) were used to sequence plasmid pRH512 according to this method. Deletions were made from the universal primer end. For making these deletions, the DNA was cut with EcoRI, filled in with α-thiophosphate dNTPs (Puteny, S. D., et al. [1981] Proc. Natl. Acad. Sci. USA 78:7350–7354) by use of the Klenow fragment of E. coli DNA polymerase, cut with Smal, and treated with exonuclease III. Samples were removed every 30 seconds, re-ligated, and used to transform E. coli SURE cells by electroporation. Sequencing reactions were carried out with the universal primer.

When a primer complementary to that sequence was prepared and used to sequence back through the RUM8 binding site (bases 3993 to 4002), the generated sequence, when translated, yielded the amino acid sequence generated from microsequencing of the 6.2 kDa CNBr polypeptide fragment.

A second sequencing method was performed using a combination of M12 shotgun sequencing with standard and universal and reverse M13 primers into M13 phage to permit single-stranded sequencing as follows. Plasmid pRH512 was sonicated to produce random fragments, repaired with bacteriophage T4 DNA polymerase, and these fragments were shotgun cloned into Smal-cut M13mp19 (GIBCO). Plaque lifts were screened with a radiolabeled probe prepared from the 4.5 kb insert found in pRH512 to identify appropriate clones for shotgun single stranded sequencing (see, e.g., Sambrook et al., supra). Sequencing of the BglII insert of pRH512 isolated it to nucleotides #0 to 4505, thus extending the sequence 5' and 3' to the spheroidin gene (FIG. 2).

Example 5

Obtaining Additional AmEPV Sequence

A DraI AmEPV DNA library was prepared by digesting genomic DNA with DraI. These DraI fragments were shotgun cloned into Smal-digested, phosphatase-treated vector M13mp19. Preparations of M13 virus and DNA were made by standard procedures (Sambrook et al., supra). Ligation and heat shock transformation procedures were performed conventionally (Sambrook et al., supra), resulting in the shotgun cloned fragments being transformed into the bacterial strain, E. coli UT481 (University of Tennessee) or the SURE strain.

Standard PCR (Innis et al. et al., supra) with 400 ng of genomic AmEPV DNA as a template was used to prepare a probe to identify a 586 bp Dral clone from nicotelleus filter replicas (plaque lifts) (Micron Separations, Inc.) of the M13 shotgun library of Dra1-cut AmEPV fragments. This was done to isolate a clone spanning a central unsequenced region of the spheroidin gene. The standard PCR primers used for this reaction were RM92 (SEQ ID NO. 15) (GCGGCTTGTTGGTAACACCTG) and RM118 (SEQ ID NO. 16) (CTGGTAGATATTCACTCG). This sequencing revealed that there was a single HindIII site at base 931 and that the 3' end of the spheroidin open reading frame (ORF) was truncated (FIG. 2).

The technique of inverse polymerase chain reaction (PCR) (Innis, M. A., et al., [1990] PCR protocol, a guide to methods and applications, Academic Press, Inc. San Diego, Calif.) was used with Clal-digested AmEPV DNA fragments which were ligated into a circle, to prepare a probe to identify clones containing a flanking sequence or to verify the absence of an intervening sequence between adjacent clones. The primers used in inverse PCR were RM82 and RM83, which were taken from the pRH512 sequence. The sequence of RM82 (SEQ ID NO. 13) was TTTCATATTAACTGGGCAACC and that of RM83 (SEQ ID NO. 14) was GGGATGGAATTGGATTGCC.

The specific PCR reaction conditions for 34 cycles were as follows: 30 seconds at 94°C for denaturation, 30 seconds at 72°C for annealing, and 1.5 minutes at 72°C for extension. Finally, the samples were incubated at 72°C to 8.5 minutes to complete extensions. The concentration of each primer was 1 µM.

The resulting 2.2 kb inverse PCR product was digested with Clal, and a 1.7 kb fragment was gel purified. The 1.7 kb PCR fragment was sequenced with RM83 as a primer. Additional PCR primers were made to the new sequence as it was identified. The sequencing process employed Sequenase, 5 pmol of each primer, and 10 to 50 ng of template. Prior to being sequenced, the PCR products were chloroform extracted and purified on spun columns (Sambrook et al., supra) of Sephacryl S-400. The DNA sequence was assembled and aligned, and consensus sequence was produced (Staden, R. [1982] Nucleic Acids Res. 10:4731–4751). Both strands were completely sequenced; the PCR product sequence was verified by conventional sequence.

The relevant Clal sites of the 1.7 kb PCR fragment are at positions 3485 and 6165. This fragment was radiolabeled and used as a probe to locate additional clones, i.e., pRH827 (307 bp), pRH85 (1.88 kb), and pRH87 (1.88 kb) from the BglII fragment library. Plasmids pRH85 and pRH87 were sequenced using the same nested exonuclease II deletions and sequencing procedure, as described above for pRH512. Sequencing of the inverse PCR products with custom-designed primers confirmed that plasmids pRH85 and pRH87 represented the same 1.88 kb BglII DNA insert in opposite orientations, but also revealed a missing 80 bp between pRH827 and pRH85. This 80 bp DNA fragment was identified in the DraI fragment, as extending from bases 4543 to 5128 cloned into M13.

The orientation of the spheroidin ORF on the physical map is shown in FIG. 1. It is interesting to note that the 1.7 kb inverse PCR fragment only hybridized to the AmEPV HindIII-G fragment. The amino acid sequence derived from the 8 and 9 kDa overlapping CnBr-generated polypeptides is found from nucleotide positions 4883 to 4957 (SEQ ID NO. 38). That derived from the 6.2 kDa polypeptide is found from nucleotides 3962 to 4012 (SEQ Q ID NO. 39), and that derived from the 15 kDa polypeptide is found from nucleotides 4628 to 4651 (SEQ ID NO. 40). Therefore, all sequences obtained from protein microsequencing were ultimately found to lie within the spheroidin ORF.
Example 6

Spheroidin Gene Transcription

The start site for spheroidin gene transcription was determined. A primer complementary to the spheroidin gene sequence beginning 65 bp downstream of the predicted initiating methionine was prepared and used for a series of primer extensions.

A. Preparation of RNA and primer extension reactions.

Six 150 mm dishes of subconfluent cells were prepared. The culture media were aspirated, and 2 ml of viral inoculum was added to each dish. The virus concentration was about 0.1 to 1 PFU per cell. The dishes were occasionally agitated during a 3 hour adsorption period. At the end of this period, the cells were rinsed with 5 ml of modified PBS. The media were replaced, and the infected cells were incubated for 72 hours at 27°C. Total RNA from the infected cells was isolated by the guanidinium thiocyanate-cesium chloride procedure (Chirgwin, J. M., et al. [1979] Biochemistry 18:5294–5299).

Primer extension reactions were carried out with primer RM165 (SEQ ID NO. 17), a 35-base oligonucleotide (GTTCCGAAAACAGATTTTTCATTTTAAAAATATC) beginning and ending 100 and 65 bp downstream, respectively, of the initiating methionine codon found in the TAAATG motif. The primer was end labeled with [γ-32P] ATP and T4 polynucleotide kinase and purified on a “spin column” (Sambrook et al., supra). For annealing, 40 μg of total infected-cell RNA and 106 cpm of radiolabeled primer were coprecipitated with ethanol. The pellet was suspended in 25 μl of hybridization buffer (80% formamide, 40 mM piperazine-N,N'-bis(2-ethanesulfonic acid) (pH 6.4), 400 mM NaCl, 1 mM EDTA (pH 8.0), denatured at 72°C for 15 minutes, and incubated at 30°C for 18 hours.

For primer extension, the RNA-primer hybrids were ethanol precipitated, resuspended, and used for five individual reactions. Each reaction contained 8 μg of total infected-cell RNA, 50 mM Tris-HCl, (pH 8.3), 50 mM KCl, 10 mM dithiothreitol, 10 mM MgCl2, 4 U of avian myeloblastosis virus reverse transcriptase (Life Sciences), 8 U of RNasin (Promega), 0.25 mM each deoxynucleoside triphosphate (dNTP), and the appropriate deoxycytosine triphosphate (dCTP), except for a control reaction, which contained no dNTP. The dNTP/ddNTP ratios were 4:1, 5:1, 5:1, and 2:1, for the C, T, A, and G reactions, respectively. The reactions were carried out at 42°C for 30 minutes.

One microliter of chase buffer (4 μl of 5 mM dNTP mixture and 1 μl of 20-U/μl reverse transcriptase) was added to each reaction mixture, which was then incubated for an additional 30 minutes at 42°C. Reaction products were separated on a sequencing gel (8% acrylamide containing 7M urea) and visualized by autoradiography. Complementarity was observed until the AAA of the upstream TAAATG motif, indicating that transcription of the gene initiates within the TAAATG element of the proposed late promoter element. Immediately upstream is a 5′ tract of noncoding poly(A) on the transcripts. The average length of the poly(A) is greater than 6 bp.

Example 7

Analysis of Spheroidin Sequence

The spheroidin ORF (G5R) was initially identified by sequencing back through the RMS8 oligonucleotide primer binding region as described above. Examination of the AmEPV spheroidin gene sequence (ORF G5R) revealed a potential ORF of 3.0 kb capable of encoding 1,003 amino acids or a protein of about 115 kDa. The ORF consists of 29% G+C, in contrast to the 18.5% reported for the entire AmEPV genome (Langridge, W. H. R., R. F. Bozarth, D. W. Roberts [1977] Virology 76:616–620). Inspection of the 92 bases upstream of the initiating ATG revealed only 7 G or C residues. Also detected was the presence of known vertebrate poxvirus regulatory sequences within the 92 bp 5′ of the spheroidin ORF. Included are three TTTT TTTT TTT gene termination signals and TAAATG, which presumably represents a late transcription start signal used to initiate transcription and translation of the spheroidin gene. Several adjacent translation termination codons are also present within the 92 bp upstream of the spheroidin ORF.

Analysis of the sequence upstream of the spheroidin gene revealed four additional potential ORFs, G1L (SEQ ID NO. 25), G2R (SEQ ID NO. 23), G3L (SEQ ID NO. 26), and G4R (SEQ ID NO. 24), discussed above. The putative amino acid sequences of these ORFs are reported in FIG. 2 (SEQ ID NO. 2, 3, 4 and 5, respectively). No significant homologies were found for the small potential polypeptides encoded by ORF G2R (SEQ ID NO. 23) or G3L (SEQ ID NO. 26). ORF G1L (SEQ ID NO. 25), however, exhibited a significant degree of homology to ORF 17 found within the HindIII-1 fragment of vaccinia virus, whose function is unknown. ORF G4R (SEQ ID NO. 24) showed homology to ORF 13 of Vaccinia virus. In vaccinia virus, the ORF 13 homolog was found very near the site of an incomplete ATG gene. The partial G6L ORF (SEQ ID NO. 27) to the right of the spheroidin gene exhibited good homology to vaccinia virus NTPase I. Much better homology (78.4% identity over 162 amino acids) was found between the partial G6L ORF (SEQ ID NO. 27) and NPH of ChEPV (Yuen, L., et al. [1991] Virology 182:403–406).

Example 8

Isolation and Sequencing of the AmEPV EcoRI-Q Fragment Containing the tk Gene

Sequencing of the EcoRI-Q fragment of genomic AmEPV of Example 1 was performed using techniques described above for spheroidin. The sequencing showed 1511 bp containing two complete and one partial ORF. Analysis of the DNA sequence of ORF Q2 (SEQ ID NO. 28) indicates the sites where the identifying degenerate oligonucleotides (RM03 SEQ ID NO. 18 and RM04 SEQ ID NO. 19) might hybridize. Two oligonucleotides, RM03 and RM04, based on different but strongly conserved regions of the tk genes of several poxviruses and vertebrates (Upton, C., et al. [1986] J. Virol. 60:920–927; Boyle, D. B., et al. [1987] Virology 156:355–365) were prepared by the methods referred to above. RM03 was the 32-fold degenerate oligonucleotide (SEQ ID NO. 18) GAT/G(C/G/A/G/A/G/A/G)/GG/G/AA/G(A/C)(G/A)T(T/C)T/GG corresponding to the amino acid residues in the vaccinia tk protein from the aspartic acid at position 82 to the phenylalanine at position 87. RM04 (SEQ ID NO. 19) was (G/GGCGG(TTT/T)/T(CG)(C/G)G/G/C/G/T/C) corresponding to 32-fold degeneracy and corresponded to the region from the glycine at position 11 to the glycine at position 16 in vaccinia. These probes were radiolabeled as described above for the RH58 probe.

The AmEPV thymidine kinase (tk) gene was identified by hybridization with the degenerate oligonucleotide probes RM03 and RM04 to a Southern blot of the EcoRI-digested EPV DNA. The EcoRI band of interest (EcoRI-Q) was isolated, purified, and ligated into a pUC18 vector (GIBCO) previously digested with EcoRI and treated with calf intestinal alkaline phosphatase. Recombinant clones were iden-
ified by the size of the insert and by hybridization to the radioactive labeled oligonucleotide probes.

One such clone was called pMEGtalk-1. The recombinant clones containing the EcoRI-Q fragment oriented in both directions relative to the pUC18 vector sequences were used for sequencing. Sequential nested deletions were generated by the method of Henikoff, cited above, as described for pRH152. These clones were used for sequencing the entire EcoRI-Q fragment.

Subsequently, these oligonucleotides and another, RM125, is a non-degenerate oligonucleotide GGTCGAAAAATCT-GATATTTC (SEQ ID NO. 20) prepared from the ORF Q1, were employed as sequencing primers to confirm their positioning as indicated in ORF Q2 (SEQ ID NO. 28). ORF Q2 potentially encodes for a protein of 182 amino acids (21.2 kDa) (SEQ ID NO. 10). ORF 03 potentially encodes a polypeptide of at least 68 amino acids but is incomplete and is transcribed in the opposite direction from ORF Q2. ORF Q1 (SEQ ID NO. 31) potentially encodes a small peptide of 66 amino acids (7.75 kDa) (SEQ ID NO. 9).

Further analysis of the EcoRI-Q fragment reveals several other points. First, the +A+T content is very high (80%). For ORF Q2, the 100 nucleotides upstream of the start codon for translation are 90% +A+T. Some potential poxvirus transcription signals were found between ORFs Q1 and Q2. The five bases immediately preceding the start codon for ORF Q1 are TAAAATG which comprise a consensus late poxvirus promoter. A potential poxvirus early transcription termination signal sequence (TTTTATT) is located 2 nt past the translation stop codon of Q2.


Example 9

Expression of the AmEpV tk Gene in a Vaccinia Virus

The AmEpV tk gene was tested functionally by cloning the gene into a vaccinia virus strain tk- mutant, as follows.

The EcoRI-Q fragment of AmEpV, described above, was inserted in both possible orientations into shuttle plasmid pHGN3.1 (Bloom, D. D., et al. [1991] J. Virol. 65:1530–1542) which had been isolated from bacterial cells by the alkaline lysis method. This EcoRI-Q DNA fragment contains the AmEpV tk open reading frame (ORF). The cloning was performed conventionally. The resulting plasmid was designated pHGN3.1/EcoRI-Q.

The plasmid was transfected by Lipofectin (GIBCO) as described specifically below into mammalian cells infected with vaccinia virus. The cells were either rat tk- human 143 tk-, or CV-1 cell lines onto which the vaccinia virus S38 was propagated. The cells were then maintained in Eagle’s Minimal Essential Medium with Earle’s salts (Massung et al. [1991] Virol. 180:347–354, incorporated by reference herein).

The VSCS vaccinia strain (Dr. Bernard Moss) contains the β-galactosidase gene driven by the vaccinia P3 promoter (P3, Lac-Z cassette) inserted into the viral tk gene. While VSCS contains an inactive tk gene due to the insertion of the β-galactosidase, portions of the vaccinia tk sequence remain. VSCS is thus tk- and, upon staining with X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside), will form blue plaques (β-galactosidase positive).

Cells were grown to 80% confluence (4x10^6 per 60 mm dish). Lipofectin solution (20 μg of Lipofectin in 50 μl of dH2O) was added to 10 μg plasmid DNA (pHGN3.1/AmEpV EcoRI-Q) in 50 μl of dH2O and incubated for 15 minutes at room temperature. After a 2 hour period of viral adsorption (m.o.i. of 2, 37°C), the monolayers were washed three times with serum-free OptiMEM. Three milliliters of serum-free OptiMEM was then added to each 60 mm dish. The Lipofectin/DNA mixture was slowly added dropwise with gentle swirling and incubated an additional 12 to 18 hours at 37°C. Fetal bovine serum was then added (10% final) and the infected cells were harvested at 48 hours postinfection.

Recombinant viruses, containing the EcoRI-Q fragment inserted into the hemagglutinin (HA) gene of vaccinia, were identified by hybridization of AmEpV EcoRI-Q fragments, radioactively labeled by procedures described above, to replicas of nitrocellulose "lifts" of virus plaques from the infected monolayer. Potential recombinants were isolated from replica filters and plaque-purified several times before testing.

The tk of AmEpV exhibits some degree of homology with the tk of vaccinia. To confirm that insertion of the AmEpV tk gene was within the HA gene of vaccinia rather than within residual tk sequences remaining in VSCS, the recombinants were examined by a series of Southern hybridizations to HindIII digests of the various viruses. When DNA from wild-type virus was hybridized to a vaccinia virus tk probe, hybridization was observed exclusively within the -3 kb HindIII-J fragment of vaccinia.

When either VSCS or either of the AmEpV tk containing recombinants was examined using the vaccinia tk probe, hybridization occurred instead to an ~8 kb fragment consistent with insertion of the -3.1 kb β-galactosidase cassette into the TK gene of the HindIII-J fragment. Hybridization of the HindIII digests of the same three viruses to radioactively labeled AmEpV EcoRI-Q DNA resulted in hybridization to a large-molecular-weight DNA fragment in the recombinant containing the AmEpV TK gene, which we have shown corresponds to the HindIII-A fragment. These results suggest that insertion of the AmEpV TK gene into vaccinia occurred within the HA gene contained within the large (~22 kb HindIII-A fragment of the virus as expected. Identical results were observed for VSCS:TKL. It is also interesting to note the lack of hybridization of the vaccinia virus (VV) and AmEpV TK probes to the heterologous poxvirus TK gene under these conditions, even though there is significant sequence homology between the vaccinia and the AmEpV TK genes.
Plaque-purified vaccinia recombinants were tested for growth on human 143 TK<sup>−</sup> cells in the presence of methotrexate. Under these conditions, only TK<sup>−</sup> virus will grow and produce plaques. Both recombinants (VSC5::AmEPV TKI and VSC8::AmEPV TKII) representing the two orientations of the AmEPV EcoRI-Q fragment gave plaques in the presence of methotrexate. All plaques from both recombinant viruses which grew in the presence of methotrexate also stained blue upon staining with X-Gal, suggesting that the TK<sup>−</sup> phenotype of these recombinants is most likely due to a functional TK gene contained within the AmEPV EcoRI-Q fragment. Since functionality is independent of the orientation of the AmEPV EcoRI-Q fragment, it is plausible that an AmEPV TK gene promoter contained within the EcoRI-Q fragment may function in vaccinia.

**Example 10**

Orientation of the TK Gene Within the AmEPV Genome

In order to determine the orientation of the Q1, Q2, Q3 ORFs within the EcoRI-Q fragment relative to the genome, a DNA primer-mediated extension reaction was performed. The EcoRI-Q fragment is contained within the PsiI-F fragment. An oligonucleotide identical in sequence to bases 182–163 (RM129, SEQ ID NO. 12) was prepared and hybridized to a heat-denatured PsiI digest of genomic AmEPV DNA. The partial hybrid was then extended with the Klenow fragment of *E. coli* DNA polymerase in the presence of radiolabeled substrates. Extension will terminate at the end of the PsiI-F fragment.

The radiolabeled product was then hybridized to an EcoRI digest of AmEPV DNA. If orientation of the gene is such that the tk ORF reads toward the end of the genome, hybridization would be expected to the EcoRI-E fragment; whereas if the gene is read toward the center of the genome, hybridization would be expected to the EcoRI-I fragment.

The results indicate hybridization not only to the EcoRI-E fragment, but also to the EcoRI-A fragment. These results infer that the orientation of the tk gene is with reading toward the left end of the genome. Hybridization of the run-off extension product also to the EcoRI-A fragment is consistent with the presence of an inverted terminal repetition, common in poxviruses, with identical sequences residing in both the EcoRI-A and the EcoRI-E fragments.

The optimal growth temperature for AmEPV in the laboratory is 28°C, whereas that of the vertebrate poxviruses is 37°C. As described herein, when the AmEPV DNA fragment containing the entire tk gene was cloned into the tk<sup>−</sup> strain of vaccinia virus, the recombinant virus was capable of growing at 37°C in the presence of methotrexate (Sigma), indicative of a tk<sup>−</sup> phenotype. This example demonstrates that the Entomopoxivirus tk gene can be successfully transferred into mammalian expression systems, and that AmEPV tk is functionally active over a considerable temperature range.

**Example 11**

Isolation and Sequencing of ChEPV and CIaEPV Spheroidin Genes

Viruses and cell culture. ChoristoneuraEPVs, ChEPV and CIaEPV, occlusion bodies (OBs) were obtained from Dr. Basil M. Arif, Forest Pest Management Institute, Sault Ste. Marie, Ontario, Canada, and can be obtained from other sources as well. The AmEPV OBs were the same as those used previously (Hall and Moyer, 1991). The gypsy moth cell line, IPLB-LD-652, used for AmEPV replication was maintained in EX-CELL 401 (JRH Biosciences, Lenexa, Kan.) supplemented with 10% fetal bovine serum. Wild type cowpox virus, Brighten red strain, was produced in CV-1 cells cultured in Eagle’s MEM with Earle’s salts.

**SDS-PAGE**

occlusion body solubilization, and total cell protein preparation. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of proteins was with a 4% acrylamide stacking gel and a 7.5% separating gel. For sample preparation, 2X Laemmli sample buffer consisting of 125 mM Tris-HCl (pH 6.8), 4% SDS (wt/vol), 10% β-mercaptoethanol (vol/vol), 20% glycerol (vol/vol), and 0.05% bromphenol blue (wt/vol) was used. Quantitation of OBs was achieved by counting a diluted suspension in a hemocytometer. The OB suspension (varying from 8x10<sup>4</sup> to 1x10<sup>6</sup> OBs/ml) usually prepared in 1X TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), was diluted 1:1 with the 2X Laemmli sample buffer and immediately boiled for 5 or 10 min. For studies designed to assay for spheroidin degradation, the OBs plus sample buffer were left at room temperature for various time periods before boiling. After electrophoretic separation, the proteins were visualized by Coomassie blue staining.

The preparation of total uninfected or infected cell protein for SDS-PAGE utilized a 2X lysis buffer consisting of 160 mM Tris-HCl (pH 9), 4% SDS (wt/vol), 4% β-mercaptoethanol (vol/vol), 10% glycerol (vol/vol), 5M urea, 20 mM EDTA (pH 8), and 0.01% bromphenol blue (wt/vol). Cells were scraped from a 150 mm diameter culture dish using a rubber policeman and centrifuged at 1000g for 5 min at 4°C. Cells were washed in appropriate cold phosphate buffered saline (PBS) and recentrifuged. Insect cells were harvested in Hank’s PBS which contained 15 g glucose per liter. The cell pellets were resuspended in a final volume of 500 μl of PBS which contained 5 μl each per ml of PMSF (phenylmethyl sulfonyl fluoride, 0.1M stock in 95% ethanol) and aprotonin (0.3 mg/ml stock). The cells were probe sonicated on ice for 15 sec., five hundred microliters of 2X lysis buffer was added, and the preparation sonicated as before. Samples were stored at −70°C and boiled before loading on an SDS-PAGE gel.

Preparation of EPV DNAs. CIaEPV DNA was obtained from Dr. Basil M. Arif, Forest Pest Management Institute, Sault Ste. Marie, Ontario, Canada. ChEPV DNA used in the polymerase chain reaction (PCR) studies was prepared directly from occlusion bodies by the agarose in situ method (Hall and Hink, 1990) and extracted by "freeze-squeeze" (Sambrook et al., 1989). AmEPV DNA was prepared from infected LD-652 cell culture by concentrating extracellular virus which was then treated with DNase I and RNase A prior to dissolution of the occlusion bodies by SDS and proteinase K digestion (Griuldi et al., 1992).

**PCR Primers and Reactions.** PCR, product purification, and dyeoxy sequencing were done as described previously (Hall and Moyer, 1991) except that the specific reaction conditions for 34 cycles were 1 min at 94°C for denaturation, 1 min at 72°C for annealing, and 2 min at 72°C for extension. Finally, samples were incubated at 72°C for 9 min to complete extensions. Typically, 50 to 100 ng of template DNA was used in the PCRs. Custom oligonucleotide primers RM82, RM83, RM92 and RM118 were described previously (Hall and Moyer, 1991). Sequences of those and other AmEPV spheroidin specific PCR primers (5's base pair numbers from Hall and Moyer, 1991) are summarized in Table 2.
### TABLE 2

<table>
<thead>
<tr>
<th>AmEPV spheroedin specific</th>
<th>Expected Size (bp)</th>
<th>Appropriate Size Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>primer pair</td>
<td>ChIEPV</td>
<td>CIEPV</td>
</tr>
<tr>
<td>12 - 9</td>
<td>314</td>
<td>-</td>
</tr>
<tr>
<td>12 - 8</td>
<td>599</td>
<td>-</td>
</tr>
<tr>
<td>12 - 6</td>
<td>828</td>
<td>-</td>
</tr>
<tr>
<td>12 - 3</td>
<td>1047</td>
<td>-</td>
</tr>
<tr>
<td>12 - 4</td>
<td>1184</td>
<td>-</td>
</tr>
<tr>
<td>1 - 4</td>
<td>290</td>
<td>+</td>
</tr>
<tr>
<td>12 - 11</td>
<td>1444</td>
<td>-</td>
</tr>
<tr>
<td>1 - 11</td>
<td>549</td>
<td>+</td>
</tr>
<tr>
<td>2 - 11</td>
<td>307</td>
<td>+</td>
</tr>
<tr>
<td>12 - 15</td>
<td>1542</td>
<td>-</td>
</tr>
<tr>
<td>1 - 15</td>
<td>648</td>
<td>+</td>
</tr>
<tr>
<td>2 - 15</td>
<td>378</td>
<td>+</td>
</tr>
<tr>
<td>12 - 13</td>
<td>1944</td>
<td>+</td>
</tr>
<tr>
<td>1 - 13</td>
<td>1050</td>
<td>+</td>
</tr>
<tr>
<td>2 - 13</td>
<td>807</td>
<td>+</td>
</tr>
<tr>
<td>5 - 13</td>
<td>632</td>
<td>-</td>
</tr>
<tr>
<td>7 - 13</td>
<td>595</td>
<td>+</td>
</tr>
<tr>
<td>14 - 13</td>
<td>426</td>
<td>+</td>
</tr>
<tr>
<td>10 - 13</td>
<td>259</td>
<td>-</td>
</tr>
</tbody>
</table>

1Primer numbers correspond to those in FIG. 5 and the sequences (5' to 3') are shown below:

- 1. RM58 - GAAGTGGACCCGGAATATGTT
- 2. RM75 - GAAAAGAATAAGATTTGGA
- 3. RM76 - AGACACTTCCAGATAATGCT
- 4. RM78 - CCGCACTTATATCCGCTC
- 5. RM79 - GTTAAACCTAAATGACC
- 6. RM82 - TTGCAATACCTGGAACCC
- 7. RM83 - GGGATGTTTATTCTTCG
- 8. RM87 - GTTGCAGTCTGATGATATCC
- 9. RM90 - TCTAGACAATAAGCATTAC
- 10. RM92 - GCCTGGTTGTGGAACATC
- 11. RM93 - CATTCTTTAGGCTAAGG
- 12. RM95 - GATCCCTTACGACAAAC
- 13. RM118 - CTCTGATATTCTACTCCCG
- 14. RM169 - ATCGCCTGATATTCTACT
- 15. RM170 - ATTACCAATGATAATGTTGC

Prime RM206 (AGATGATGATGAAAATGTTGGA) (SEQ ID NO. 50) was from bases 2379 through 2398 and RM212 (GATAATAGATCCTCGGTTCG) (SEQ ID NO. 51) from bases 2077 through 2096 of the ChIEPV NPH I sequence (Yuen et al., 1991).

Cloning and double strand plasmid sequencing. BglII clones in both orientations (1.06 kb and 1.78 kb) covering the unsequenced 5' end of the AmEPV NPH I gene were selected from an AmEPV BglII fragment library (Hall and Moyer, 1991) by hybridization with an AmEPV 13 kb HindIII fragment probe. Plasmids were cloned in the Escherichia coli SURE strain (Stratagen, La Jolla, Calif.). Plasmids were sequenced by use of exonuclease III deletions and dye deoxy sequencing as described (Hall and Moyer, 1991). Both strands were completely sequenced.

Radiolabeled probes, Southern blotting, and hybridization. Random oligolabeling of DNA for probes, Southern transfer and hybridization (at 65°C with BLOTTO) were as described previously (Hall and Moyer, 1991).

Protein microsequencing. Lys-c endopeptidase digestion of the 115 kDa protein of ChIEPV OBs purified and recovered from SDS-PAGE gels was used to generate internal peptide fragments for sequencing on a gas phase sequencer.

Antibody preparation and Immunoblotting. A preparation of total occlusion body antigens was prepared by solubilizing purified AmEPV occlusion bodies purified from infected cell cultures. Rabbits were then intradermally injected with 100–200 μg of antigen per rabbit in Freund’s complete adjuvant. One month later, the rabbits were boosted with about 500 μg each of the same antigen in incomplete adjuvant. After an additional 3 weeks, the rabbits were boosted with 200 μg of antigen in incomplete adjuvant.

Eleven days later, the immune serum was collected. This serum is referred to as occlusion body antiserum. Monospecific spheroedin antibodies were prepared from this serum based on immunofinity of individual antibodies for purified spheroedin (Harlow and Lane, 1988). Samples of the sera were adsorbed to a nitrocellulose blot prepared from a preparative SDS-PAGE gel of solubilized AmEPV occlusion bodies. The section of the blot containing the 115 kDa AmEPV spheroedin and bound antibodies was then excised. The monospecific AmEPV spheroedin antibodies were eluted using 100 mM glycine, pH 2.5. After neutralization and dilution, the monospecific spheroedin antibodies were used to probe Western blots.

For immunoblotting, duplicate samples run on SDS-PAGE gels were prepared. One-half of the gel was stained with Coomassie blue, and the other half (containing prestained molecular mass markers) was transferred to nitrocellulose membrane in Tris-glycine buffer (Harlow and Lane, 1988) using the BioRad Trans Blot at 250 mA for 3 hours. The blot was blocked for 2 hr using BLOTTO (0.5% nonfat dry milk in TBS: 0.01M Tris, pH 8, 0.15M NaCl). Dilutions of the antibody were prepared in BLOTTO and adsorbed to the blot overnight. The blot was washed 3 times in TBS at room temperature, and the secondary antibody (goat anti-rabbit conjugated to alkaline phosphatase) at a 1:1000 dilution was adsorbed to the blot for 1/2h. The blot was washed with TBS as previously. Secondary antibody reactions and color development was as described (Harlow and Lane, 1988).

SDS-PAGE of solubilized occlusion bodies. When purified occlusion bodies of ChIEPV, CIEPV, and AmEPV are solubilized in Laemmli sample buffer, boiled immediately, and analyzed on SDS-PAGE gels, Coomassie blue staining shows the major protein to be about 115 kDa in each case. The Choristoneura EPVs show a second 47 kDa protein in lesser amounts. Other minor proteins were also observed.

The purported ChIEPV spheroedin gene has a coding capacity of 47 kDa (Yuen et al., 1990) despite the fact that the observed size of the corresponding spheroedin appears to be 115 kDa, which is similar to that observed for the AmEPV spheroedin. This discrepancy has been explained by suggesting that the ChIEPV spheroedin exists as relatively unstable dimers, which dissociate under a variety of conditions to monomers of 47 kDa (Yuen et al., 1990). The AmEPV spheroedin of 115 kDa, however, shows no such propensity for dissociation.

Prior to dissolution, occlusion bodies are stable and routinely stored at 4°C in buffer. The only discernible difference in methods of occlusion body solubilization and preparation is the incubation time in SDS buffer before boiling. The relative instability of the various spheroedins was evaluated by incubating the occlusion bodies of all three EPVs at room temperature for up to two hours in SDS solubilizing buffer before boiling the samples. While some degradation of the 115 kDa protein was observed for the CIEPV OB preparation, little if any degradation of the ChIEPV or AmEPV preparations was observed. The CIEPV
115 kDa protein was degraded to a variety of smaller proteins but not in a fashion suggesting a relationship to the 47 kDa protein. Whether the OB suspension was in 1X TE (10 mM Tris, 1 mM EDTA, pH 8) or in deionized water prior to dissolution has no effect on the subsequent degradation CIEPV occlusion bodies.

Discovery of an AmEPV spheroidin gene homolog in ChbEPV and CIEPV. Selected oligonucleotide primers derived from within various regions of the AmEPV spheroidin gene were selected as appropriate primer pairs for PCR to look for the spheroidin gene in ChbEPV and CIEPV. The relative positions of these primers within the spheroidin gene are shown in FIG. 5. Table 2 lists the specific primer pairs and sequences, the expected size of the PCR products based on the AmEPV spheroidin sequences, and the results when these primers were used in conjunction with ChbEPV or CIEPV templates. From Table 2, for ChbEPV template 10 out of 19 primer pairs resulted in an appropriate size product expected if the two genes were similar. For CIEPV this was 13 out of 19 primer pairs. We chose one of the products (1 kb) generated from CIEPV DNA by primer pair 1–13 (in Table 2) for further analyses.

This PCR product was radiolabeled and used as a probe for a blot containing restriction fragments of both CIEPV and AmEPV DNAs. All hybridizations to CIEPV showed predominant, specific hybridization signals. The CIEPV derived probe also shows appropriate, discrete hybridizations to AmEPV DNA; i.e., a 13 kb HindIII fragment, a 20 kb EcoRI fragment, a 4.5 kb BglII fragment, and a 4.5 kb BstBI fragment. This pattern of hybridization to AmEPV is that expected for hybridization to the AmEPV spheroidin gene.

Partial sequence of a spheroidin-like gene in ChbEPV or CIEPV. Further indications of the existence for an AmEPV spheroidin gene homolog in the genome Choristoneura EPVs come from PCR product sequencing of the 1 kb Choristoneura PCR products (primer pair 1–13; Table 2), The resulting sequences derived from the ChbEPV or CIEPV DNA with a comparison to the AmEPV sequence is shown in FIG. 6A. When the deduced amino acid sequence of this region is compared (FIG. 6B), a very high degree of homology is found between all three viruses.

The spheroidin-like gene in ChbEPV and CIEPV is expressed. Samples of the ChbEPV =115 kDa protein were isolated from SDS-polyacrylamide gels, and treated with lys-c endopeptidase to generate peptides for microsequencing. Three of the resulting peptides were analyzed, and the amino acid sequence was compared to the spheroidin of AmEPV. The ChbEPV sequences obtained were homologous to three corresponding regions of the AmEPV spheroidin (FIG. 6C). These results demonstrate that the Choristoneura viruses not only contain a spheroidin-like gene, but that gene is expressed to yield a polypeptide within occlusion bodies of similar size and sequence to the previously-identified spheroidin protein of AmEPV.

We have also addressed the question of whether the AmEPV spheroidin homolog in ChbEPV and CIEPV is expressed by a Western blot analysis of the proteins of ChbEPV and CIEPV using antibody derived against either AmEPV occlusion bodies or monospecific sera against AmEPV spheroidin. Sera directed against purified AmEPV occlusion bodies recognize proteins of â115 kDa in both ChbEPV and AmEPV. Stronger signals are observed in the AmEPV samples as expected. An AmEPV protein of 38 kDa is also recognized in the samples. Weak binding is also observed to the abundantly expressed protein of the vertebrate cowpox virus which was used as a control. When immunoaffinity purified, monospecific AmEPV spheroidin sera is used, the 115 kDa protein of ChbEPV also cross-reacts. Similar results were obtained with CIEPV. These results also support the conclusion that the two Choristoneura viruses and AmEPV encode a very similar major occlusion body protein of 115 kDa which in AmEPV corresponds to the spheroidin gene.

Co-linearity of AmEPV, ChbEPV, and CIEPV maps in the spheroidin region. We have shown that the gene adjacent to the 3' terminus of the AmEPV spheroidin gene is NPH I (NTPas I) in polarity opposite to that of the spheroidin gene. A NPH I gene from ChbEPV has been sequenced (Yuen et al., 1991).

The sequence of the NPH I (or NTPas I) gene of AmEPV is presented in FIG. 7. The 3' end sequence of this gene is provided in FIG. 1 and the nucleotide numbering system depicted in FIG. 7 results from appending the 5' end of the sequence to the sequence provided in FIG. 1.

When the complete coding sequence and deduced amino acid sequence of the AmEPV NPH I gene was compared to the already published ChbEPV NPH I gene (Yuen et al. 1991), the two genes show 89% amino acid and 86% nucleotide identity. Both proteins have a deduced 648 amino acids, with the major difference being that AmEPV has deleted amino acid number 127 and has one extra amino acid at the very end of the sequence. Both genes show the typical poxvirus late gene promoter sequence motif, TAAATG, at the beginning of the open reading frame as well as the A+T rich sequence upstream of the gene. Of the 30 bases preceding the starting ATG, only a single G differentiates the AmEPV and ChbEPV. The intergenic region between the spheroidin and NPH I genes begins to diverge immediately following the NPH I open reading frame (ORF) at the downstream 3' end.

The proximity of the spheroidin and NPH I genes in both Choristoneura viruses was tested by performing a series of PCR reactions using the PCR primers shown in FIG. 8 which were designed based on the AmEPV gene arrangement and were used with either ChbEPV or ChbEPV. Primers RM206 and RM212 are described above in this example, RM83 and RM161 in Table 2. Primer pair RM161-RM212 failed to give a PCR product with ChbEPV DNA indicating sequence differences between ChbEPV and CIEPV in the RM161 primer binding region. However, the other three reactions generated PCR products ranging in size from â1-2.5 kb, indicating that the NPH I and spheroidin genes are adjacent and arranged similarly to AmEPV. Based on the PCR products, the intergenic distance between the Choristoneura spheroidin and NPH I genes appears to be at least about 450 bp longer than the same region in AmEPV. The increased size of this intergenic region is large enough to perhaps allow for a small ORF to be present between the two genes in the Choristoneura viruses. However, PCR product sequencing of the RM83 - RM212 and the RM161 - RM206 PCR products using RM206 as the primer failed to show any ORF of significant size.

Our results suggest that the published Choristoneura EPV gene identified by Yuen et al. (1990) as the spheroidin gene, is incorrect. Our evidence shows instead that the spheroidin gene of Choristoneura EPVs is virtually identical to that of AmEPV (FIG. 6), i.e., encodes a 115 kDa protein and is expressed. Hence this protein is highly conserved amongst all three viruses.
The AmEPV spheroidin gene found in the Choristoneura EPV genomes and the NPH I genes were found to be immediately adjacent to each other in all three viruses. Although the intergenic distance between the two genes is somewhat different between the two Choristoneura viruses and AmEPV, it would appear that the genes in this region of the viral DNA are co-linear.

We have shown that the linear conserved core of genes found in vertebrate poxviruses is not maintained in AmEPV. Based on the data presented here, it appears that the entomopoxviruses have evolved and maintained a common core of co-linear genes, different from their vertebrate counterparts.

It should be understood that the examples and embodiments described herein are for illustrative purposes only. Various modifications or changes in light thereof will be suggested to persons skilled in the art by this specification. The subject invention encompasses recombinant polynucleotide sequences, plasmids, vectors, and transformed hosts which are equivalent to those which are specifically exemplified herein in that the characteristic expression features are retained in said equivalent constructs even if inconsequential modifications to the DNA sequence have been made. For example, it is within the skill of a person trained in the art to use a fragment of the spheroidin gene’s non-coding region which is upstream of the structural gene in order to achieve the desired level of expression. Such fragments of the regulatory sequences fall within the scope of the current invention, so long as the desired level of expression which is characteristic of this system is retained. Furthermore, inconsequential changes to the nucleotide sequences can be made without affecting the disclosed functions of these sequences. Such modifications also fall within the scope of the current invention and are to be included within the spirit and purview of this application and the scope of the appended claims.

**SEQUENCE LISTING**

1. **GENERAL INFORMATION:**
   
   (iii) NUMBER OF SEQUENCES: 66

2. **INFORMATION FOR SEQ ID NO:1:**
   
   (i) **SEQUENCE CHARACTERISTICS:**
   
   (A) LENGTH: 9457 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: double
   (D) TOPOLOGY: unknown

   (ii) **MOLECULE TYPE:** DNA (genomic)

   (vi) **ORIGINAL SOURCE:**
   
   (A) ORGANISM: Amaeta moorei entomopoxivirus

   (ix) **FEATURE:**
   
   (A) NAME/KEY: CDS
   (B) LOCATION: complement (65..1459)

   (ix) **FEATURE:**
   
   (A) NAME/KEY: CDS
   (B) LOCATION: 1476..2151

   (ix) **FEATURE:**
   
   (A) NAME/KEY: CDS
   (B) LOCATION: complement (2239..2475)

   (ix) **FEATURE:**
   
   (A) NAME/KEY: CDS
   (B) LOCATION: 2502..2987

   (ix) **FEATURE:**
   
   (A) NAME/KEY: CDS
   (B) LOCATION: 3086..5091

   (ix) **FEATURE:**
   
   (A) NAME/KEY: CDS
   (B) LOCATION: complement (6277..6768)

   (x) **SEQUENCE DESCRIPTION: SEQ ID NO:1:**

   AGATCTGATG TCTATATAT ATGACAAATTT GTATAATTTA ATGATATT TAAAATCTAA 60
   GATATATTAG ATATTATTC TACTAAATCT TCTCATTATG AATATAACTA TCTAATCT 120
   TTTTTTATT ACTACATACA TCTATAATTC TATATTGATT TTTTTTATAC ATATCATATA 180
   ATCCCATAAA CTTTTTATTT TTTATAATTA AATATTCTAA TGTATTTTTA AATTGTCAGA 240
TAC TAT TCTA GAAATAAATA ATGCACCTCT ATAACTACTA GCCAATAAT 390
CACCATAAAA ACTCATAGAA TAAATATTT TTTTAAATTC AAAAAATAGT TTTAGTTGA 360
AATAAATCT ATATAATAAAA AATTATATAT TAAACATACC ACAATCGGGA CATCATATT 420
GATAATCCAA AGTAATTTAA AAGTAATATT TTAACATTTTT AAAATATCA TTTAAATT 480
CTGATAGTAC ATCAATGAT AAAAAAGCAT ATTAGTATTT AGAGATCTA TTGTAAGTTT 540
TATGGCTTTTT TATAGTCATA TCGAGTTCAA AAAAAACATAA TTATTTATT TGTATTATAA 600
GTCTGAGTAT ATAACACTAA CTAATTTTTA AGTATGACGC TTTTTTTCT TTTATCAAAGT 660
GTTTATATCT TACGCCAACA GTAAAAATGAT CATTATAAAT TATAGGAAAC ATAAAAAAC 720
TTTTTTATAT ATCTATTAAAA AAAATTTTTT CTCTATCTTC AAGTTATAG CATCTCATAG 780
ATGAGCTATC TGATGCAATA TTTTTATCAG TTTTTTCAAA TAAAATCAA TGAAAAATAAT 840
CATAATCTGT ATAAATCATG GTAATAAACAT ATATACAAAT CCGAAACCTTA 900
ACCACCTAGA TTTATCATGT TTTCTGGGT AAGCTTAAAG TTAGAGTATTA AATCCCAAAG 960
GGGATTCTTC ATTTGGACCA TCCAAATCAT CATAAAATGT GCAAATGTA GAAAAACTTC 1020
TTGTATTGGA TATYCTTGAT TTTGAGAAAG ACTTCTTCTAT ATATACATG GAATGCCCTT 1080
TTATATTCTT AGATGATAA AAAAGTATTTA TTTATTTT ATTTATTTTT ATATATTTTTT 1140
TTATAGATCA AATAGAAGAA AATTTTTTTT TAATATTATT TTTCAATAAT ATATTAAAAA 1200
TATGATTGGG ATCTCAACCT ATATCATTAT ATATATTTTT CAAGTATTTTT ATAGGTAAAA 1260
ATGTACCATT ACCTTTGTTT TCATCATCAT CATCTATTTT TTTCAATAAT GCTATATTAG 1320
CATTAGATTAT ATATTTAATA GGGATTTAAA AATATACCAT ATTATCTATT TTTACTAAAA 1380
ATACATAGA CATATAATTA ATACACAGATT CTGCGCATTTT TAAATTTTTA TTTGGAAACT 1440
TTCAATTATTT ATATTCATT TTTATTTTTA TAA ATG TTT CTA GTT TAT TAC AAT Mot Phe Leu Val Tyr Phe Asn 1 1494
ACA TTT TTA ATA ATA ATT TTA TTA TTA GGT ATT ATA GGT ATT TAT ATA Thr Phe Leu Ile Ile Ile Leu Leu Phe Gly Ile Ile Gly Ile Tyr Ile 10 1 5 15 20
TTA ACA TTT GTG TTT TAT ATA GAT TTT TTA ATA AAT AAT AAT AAA ATA Leu Thr Phe Val Phe Asn Ile Asp Phe Leu Ile Asn Asn Asn Lys Ile 25 30 35
TAT ATA TTA TCA TAT AAC GCA ACT AAT ATA AAC AAT ATA AAT TTA Tyr Ile Leu Ser Tyr Asn Ala Thr Asn Asn Asn Asn Asn Leu 40 45 50 55
AAT TTA TAC GAT TAT TCA GAT ATT ATA TTT TTT ACA AAT TTT AAC ATA Asn Leu Tyr Asp Tyr Ser Asp Ile Ile Phe Leu Thr Asn Phe Asn Ile 60 65 70
AAT AAT AAT CTT TTA GTA ACA CAA GCT AAT AAT TTA CAA GAT ATA CCA Asn Asn Asn Leu Leu Leu Val Thr Glu Ala Asn Asn Glu Asn Asn Ile Pro 75 80 85
ATA TTT AAT GTA AAT ATT ATA TCT AAT CAA TAT ATT TAT TAT TAT GTA Ile Phe Asn Val Asn Ile Ile Ser Asn Glu Tyr Asn Phe Tyr Ser 90 95 100 105 110
GCG TCT AGT AAT AAT GTA AAT ATA TTA TTA GGA TTA AGA AAA ACA TTA Ala Ser Ser Asn Asn Val Asn Ile Leu Leu Gly Leu Arg Lys Thr Leu 120 125 130 135 141
AAT ATA AAT AGA AAT CCA TTT TTA TTA TTA TTA TTA TTA AAT ACA TCT CTA GCT Asn Ile Asn Arg Asn Pro Phe Leu Leu Phe Arg Asn Thr Ser Leu Ala 148 153 159
ATA GTT TTC AAT AAT GAA ACT TTT CAC TOT TAT AAT AGT TCA AAT Ile Val Phe Asn Asn Glu Thr Phe His Cys Tyr Ile Ser Ser Asn 1926
<table>
<thead>
<tr>
<th>140</th>
<th>145</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAA AAT AGT GAT GTA TTA GAT ATA GTA TCA CAT ATA GAA TTT ATG AAA</td>
<td>1974</td>
<td></td>
</tr>
<tr>
<td>Gln Asn Ser Asp Val Leu Asp Ile Val Ser His Ile Glu Phe Met Lys</td>
<td>160</td>
<td>165</td>
</tr>
<tr>
<td>TCT AGA TAT AAT AAA TAT GTA ATT GTA AGA GAA ATA CCC GTA AAT AAT</td>
<td>2022</td>
<td></td>
</tr>
<tr>
<td>Ser Arg Tyr Asn Lys Tyr Val Ile Ile G1y Glu Ile Pro Val Asn Asn</td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>AAT ATA TCT ATT AAT ATA TTA AAT AAT TTT GCT ATT ATA ACT AAT</td>
<td>2070</td>
<td></td>
</tr>
<tr>
<td>Asn Ile Ser Ile Asn Asn Ile Leu Asn Asn Phe Ala Ile Ile Thr Asn</td>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>GTG AGA TTA ATA GAT AAA TAT AAC TCT ATA ATA TCA TTT TTA AAT ATC</td>
<td>2118</td>
<td></td>
</tr>
<tr>
<td>Val Arg Leu Ile Asp Lys Tyr Asn Ser Ile Ser Phe Leu Asn Ile</td>
<td>200</td>
<td>205</td>
</tr>
<tr>
<td>AAC GTA GGA ACA CTT TTT GTC ATA AAT CCA TAAATATTAG TAAATACAC</td>
<td>2168</td>
<td></td>
</tr>
<tr>
<td>Asn Val Gly Thr Leu Phe Val Ile Asn Pro</td>
<td>220</td>
<td>225</td>
</tr>
<tr>
<td>TAA CAAAATT TTT ATTAAAAA TGAATAAAAA ATATATTGGT ATGGTCAATA TTTTTATCA</td>
<td>2228</td>
<td></td>
</tr>
<tr>
<td>TTTTACAGTC TTTTTTTTT TTTTTGTTT TAGGTTAAT TTTACTTCTT AAAAGTTTAT</td>
<td>2288</td>
<td></td>
</tr>
<tr>
<td>CTCCCCAAAAC ATCTACAGTA GATGGTTTAT TAGATTTCTGTT GTATACACA TCTGCTGAT</td>
<td>2348</td>
<td></td>
</tr>
<tr>
<td>TTGCCGGCATT TGTATCCAAA CCATAATATC CAGGTCCTATA ATATTCTTTA AAAACTYGGG</td>
<td>2408</td>
<td></td>
</tr>
<tr>
<td>ATGGAGATAC TCTTCAGTT TTTAATATAT TAAAATACCC AAGATATTAT TTTTTGATG</td>
<td>2468</td>
<td></td>
</tr>
<tr>
<td>AAGACATAAT TGATATTTATA ATACTTTTATA GAT ATG TCA ATA TTT ATC TAG TAT</td>
<td>2532</td>
<td></td>
</tr>
<tr>
<td>Met Ser Ile Phe Ile Tyr Tyr</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>ATT TTC AAC AAT AGA TTT TAT ATA TAT AAA AAG AAG ATG AAT ACT GTA CAA</td>
<td>2570</td>
<td></td>
</tr>
<tr>
<td>Ile Phe Asn Asn Arg Phe Tyr Ile Tyr Lys Met Ser Thr Val Glu</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>ATT TTA GTT GTC ATA TTA ATA ACA ACA GCA TTA TCT TTT CTA GTT TTT</td>
<td>2618</td>
<td></td>
</tr>
<tr>
<td>Ile Leu Val Val Ile Leu Ile Thr Thr Ala Leu Ser Phe Leu Val Phe</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>CAA TTA TGG TAT TAT GCC GAA AAT TAC GAA TAT ATA TTA AGA TAT AAT</td>
<td>2666</td>
<td></td>
</tr>
<tr>
<td>Gln Leu Thr Tyr Tyr Ala Glu Asn Tyr Glu Tyr Ile Leu Arg Tyr Asn</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>GAT ACA TAT TCA AAT TTA CAA TTT GCG AGA AAC GAT AAA TAA TAT TTT</td>
<td>2714</td>
<td></td>
</tr>
<tr>
<td>Asp Thr Tyr Ser Asn Leu Glu Phe Ala Arg Ser Ala Asn Ile Asn Phe</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td>GAT GAT TTA ACT GTT TTT GAT CCC AAC GAT AAA GTT TTT AAT GTT GAA</td>
<td>2762</td>
<td></td>
</tr>
<tr>
<td>Asp Asp Leu Thr Val Val Asp Pro Asp Pro Asn Asp Phe Val Glu</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>GAA AAA TGG CGC TGT GCT TCA ACT AAT AAT AAT ATA TTT TAT GCA GTT</td>
<td>2810</td>
<td></td>
</tr>
<tr>
<td>Glu Lys Thr Arg Cys Ala Ser Thr Asn Asn Ile Phe Tyr Ala Val</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>TCA ACT TTT GGA TTT TTA AGT ACA GAA AGT ACT GTT ATT AAT TTA ACA</td>
<td>2858</td>
<td></td>
</tr>
<tr>
<td>Ser Thr Phe Gly Phe Leu Ser Thr Glu Ser Thr Gly Asn Leu Asn Thr</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>TAT ACA AAT TCT AGA GAT TGT ATT ATA GAT TTA TTT TCT AGA ATT ATA</td>
<td>2906</td>
<td></td>
</tr>
<tr>
<td>Tyr Thr Asn Ser Arg Asp Cys Ile Leu Asp Leu Phe Ser Arg Ile Ile</td>
<td>120</td>
<td>125</td>
</tr>
<tr>
<td>AAA ATA GTA TAT GAT CCT TGT ACT GTC GAA ACA TCT AAC GAT TGT AGA</td>
<td>2954</td>
<td></td>
</tr>
<tr>
<td>Lys Ile Val Tyr Asp Pro Cys Thr Val Glu Thr Ser Asn Asp Cys Arg</td>
<td>140</td>
<td>145</td>
</tr>
<tr>
<td>TTA TTA AGA TTA TTT GAG GCC AAA GCA TCA TAAATACATT ATAAATATTAT</td>
<td>3004</td>
<td></td>
</tr>
<tr>
<td>Leu Leu Arg Leu Leu Met Ala Asn Thr Ser</td>
<td>155</td>
<td>160</td>
</tr>
<tr>
<td>TAAAATATCA ATCAATATTT TTATATATAT TTATCTAAA AGGACTTTTTT ATTTTTTATA</td>
<td>3064</td>
<td></td>
</tr>
<tr>
<td>TATTAATATT AAAAA ATG AGT AAC GTA CCT TTA GCA ACC AAA ACA ATA AGA</td>
<td>3115</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Met Ser Asn Val Pro Leu Ala Thr Lys Thr Ile Arg</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>AAA</td>
<td>TTA</td>
<td>TCA</td>
</tr>
<tr>
<td>Lys Leu Ser Asn Arg Lys Tyr Glu Ile Lys Ile Tyr Leu Lys Asp Glu</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>AAT</td>
<td>ACT</td>
<td>TGT</td>
</tr>
<tr>
<td>Asn Thr Cys Phe Glu Arg Val Val Asp Met Val Val Pro Leu Tyr Asp</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>GTG</td>
<td>TGT</td>
<td>AAT</td>
</tr>
<tr>
<td>ATA</td>
<td>GAA</td>
<td>GTA</td>
</tr>
<tr>
<td>GGC</td>
<td>GAT</td>
<td>ACA</td>
</tr>
<tr>
<td>GTA</td>
<td>AAC</td>
<td>GAA</td>
</tr>
<tr>
<td>AAT</td>
<td>GTA</td>
<td>TCA</td>
</tr>
<tr>
<td>ACT</td>
<td>CTT</td>
<td>AAT</td>
</tr>
<tr>
<td>TTT</td>
<td>ATT</td>
<td>TAT</td>
</tr>
<tr>
<td>GAA</td>
<td>ATT</td>
<td>CAA</td>
</tr>
<tr>
<td>ATT</td>
<td>ACT</td>
<td>ATA</td>
</tr>
<tr>
<td>TAT</td>
<td>TTA</td>
<td>GAA</td>
</tr>
<tr>
<td>ATT</td>
<td>ACT</td>
<td>TTA</td>
</tr>
<tr>
<td>GAT</td>
<td>ATA</td>
<td>TAT</td>
</tr>
<tr>
<td>GAT</td>
<td>TTA</td>
<td>CTT</td>
</tr>
<tr>
<td>GGA</td>
<td>TTA</td>
<td>ATT</td>
</tr>
<tr>
<td>ATT</td>
<td>GTA</td>
<td>GAA</td>
</tr>
<tr>
<td>AAG</td>
<td>GAT</td>
<td>GAA</td>
</tr>
<tr>
<td>ACA</td>
<td>GAA</td>
<td>GTA</td>
</tr>
</tbody>
</table>

| ACG | CTC | AGA | GTA | TAT | AAA | AAG | TTT | GAT | AAA | TCT | CAT | TTA | AAA | ATT | GTA |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Thr | Leu | Arg | Val | Tyr | Lys | Lys | Phe | Asp | Lys | Ser | His | Leu | Lys | Ile | Val |
| 320 | 325 | 330 |
| ATG | CAT | AAC | AGA | GGA | AGT | GGT | AAT | GTA | TTT | CCC | TTA | AGA | TCA | TTA | TAT |
| Met | His | Asn | Arg | Gly | Ser | Gly | Asn | Val | Phe | Pro | Leu | Arg | Ser | Leu | Tyr |
| 335 | 340 | 345 |
| CTC | GAA | TGG | TCT | ACT | GAT | TTA | AAA | GGA | TAT | CCA | GTT | AAA | GCA | TCT | GAT |
| Leu | Glu | Leu | Ser | Asn | Val | Thr | Asp | Val | Gly | Ser | Val | Pro | Val | Leu | Asp | Thr |
| 355 | 360 |
| TCG | AGA | TTA | GAT | GTT | GGT | ATT | TAC | AAA | TTA | AAT | AAA | ATT | TAT | GTA | GAT |
| Ser | Arg | Leu | Asp | Val | Gly | Ile | Tyr | Lys | Leu | Asn | Lys | Ile | Tyr | Val | Asp |
| 365 | 370 | 375 | 380 |
| AAC | GAC | GAA | AAT | AAA | ATT | ATA | TTG | GAA | GAA | ATT | GAA | GCA | GAA | TAT | AGA |
| Asn | Asp | Glu | Asn | Lys | Ile | Ile | Leu | Glu | Ile | Gln | Ala | Gln | Tyr | Arg |
| 385 | 390 | 395 |
| TGC | GGA | AGA | CAA | GTA | TTC | CAC | GAA | CGT | GAT | AAA | CTT | AAT | AAA | CAC | CAA |
| Cys | Gly | Arg | Val | Pro | Gly | Asp | Leu | Arg | Val | Lys | Leu | Asn | Lys | Pro | His |
| 400 | 405 | 410 |
| TTT | AAA | TAT | ACT | CCC | AAA | TGT | CCA | TCC | AAA | ATT | GAT | TTA | AAC | ACG | ACA |
| Cys | Lys | Tyr | Thr | Pro | Lys | Gln | Pro | Phe | Glu | Val | Pro | Val | Asn | Ser | Pro |
| 415 | 420 | 425 |
| GAT | ACT | AGC | ATT | CAC | TTA | TAT | GAT | ATT | TCT | AAT | GAT | TTT | TGG | 77A | AAA | CCT |
| Asp | Thr | Thr | Ile | His | Leu | Tyr | Ile | Ser | Gly | Ile | Ser | Asp | Val | Leu | Lys | Pro |
| 430 | 435 | 440 |
| AAA | GTA | CCC | AAA | AAT | TTA | AGA | CTT | TGG | GGA | TOG | ATT | TTA | GAT | TGC | GAT |
| Lys | Val | Pro | Lys | Asn | Leu | Arg | Leu | Thr | Gly | Thr | Ile | Leu | Asp | Cys | Asp |
| 445 | 450 | 455 | 460 |
| ACT | TCT | AGA | ATT | AAA | CAT | ATG | GCT | GAT | GGA | TCT | GAT | TTA | GAT | TGC | GAT |
| Thr | Ser | Arg | Phe | Ile | His | Met | Ala | Asp | Gly | Ser | Leu | Asp | Leu | Ser | Asp |
| 465 | 470 | 475 |
| CTT | GAC | GTT | AGG | CTT | AAT | AGA | GAT | ATA | TGT | TTA | AAA | CAA | GCC | ATA |
| Leu | Asp | Val | Arg | Leu | Asn | Arg | Asp | Ile | Cys | Leu | Lys | Gln | Ala | Ile |
| 480 | 485 | 490 |
| AAA | CAA | CAT | TAT | ACT | AAT | GTA | ATT | ATA | TTA | GAG | TAC | GCA | AAT | ACA | TAT |
| Lys | Gln | His | Tyr | Thr | Leu | Arg | Val | Ile | Ile | Leu | Tyr | Arg | Asn | Thr | Tyr |
| 495 | 500 | 505 |
| CCA | AAT | TGC | ACA | TTA | TCA | TGG | GAT | AAT | AAT | AGA | TTT | AAT | GTA | TTT |
| Pro | Asp | Cys | Thr | Leu | Ser | Leu | Gly | Asn | Asp | Arg | Phe | Asn | Val | Phe |
| 510 | 515 | 520 |
| GAT | ATG | AAT | GAT | AAC | AAA | ACT | ATA | TCT | GAG | TAT | ACT | AAC | TTT | ACA | AAA |
| Asp | Met | Asp | Asp | Lys | Thr | Ile | Ser | Gly | Ser | Thr | Phe | Thr | Asp | Thr | Lys |
| 525 | 530 | 535 | 540 |
| AGT | AGA | CAA | GAC | CTT | AAT | AAC | ATG | TCA | TGT | ATA | TTA | GCA | ATA | AAC | ATA |
| Ser | Arg | Gln | Asp | Leu | Asn | Met | Ser | Cys | Ile | Leu | Gln | Ile | Asn | Ile |
| 545 | 550 | 555 |
| GGT | AAT | TCC | GTA | AAT | ATT | AGT | TTT | CCT | GGT | TGG | GTA | ACA | CCT | CAC |
| Gly | Asn | Ser | Val | Asn | Ile | Ser | Ser | Ser | Pro | Gly | Thr | Val | Thr | Pro | His |
| 560 | 565 | 570 |
| GAA | GCT | AAA | ATT | CTA | AGA | TCT | GGT | TGT | GCT | AGA | GTT | AGA | TTT | TGT |
| Gln | Ala | Lys | Ile | Leu | Arg | Ser | Gly | Cys | Ala | Arg | Val | Gln | Phe | Cys |
| 575 | 580 | 585 |
| AAA | TCA | TCT | TGT | GAT | CTT | CCT | AAT | AAG | AGA | TCC | TCT | CAC | TAT | GCT | GCT |
| Lys | Ser | Phe | Cys | Asp | Cys | Thr | Ser | Leu | Arg | Phe | Leu | Arg | Asn | Thr | Ala |
| 590 | 595 | 600 |
| GAT | TCT | GTA | AGT | TTA | CTA | TTT | ATG | GCT | AGA | TGT | CTA | TAT | GPT | ATT | GAT |
| Asp | Leu | Val | Ser | Leu | Leu | Pro | Cys | Asn | Tyr | Val | Asn | Ile | Gly | Ile |
| 605 | 610 | 615 | 620 |
| AAC | GAA | GCA | GTA | TGC | GAA | TAT | CCT | GGA | TAT | GTC | ATA | TTA | TTC | GCA |
| Asn | Glu | Ala | Val | Cys | Glu | Tyr | Pro | Gly | Tyr | Val | Ile | Leu | Phe | Ala | Arg |
| 625 | 630 | 635 | 640 |
| GCT | ATT | AAA | GTA | ATT | AAT | GAT | TTA | TTA | TTA | ATT | AAC | GGA | GTA | GAT | AAT |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ala | Lle | Lys | Val | Ile | Asn | Asp | Leu | Leu | Leu | Ile | Asn | Gly | Val | Asp | Asn |
| 640 | 645 |
| CTA | GCA | GGA | TAT | TCA | ATT | TCC | TTA | CCT | ATA | CAT | TAT | GGA | TCT | ACT | GAA |
| Leu | Ala | Gly | Tyr | Ser | Ile | Ser | Leu | Pro | Ile | His | Tyr | Gly | Ser | Thr | Glu |
| 655 | 660 | 665 |
| AAG | ACT | CCA | CAA | AAT | GAA | GAA | TAT | GGT | GGT | GGT | CAT | GAT | AAA | TTT | AAA |
| Lys | Thr | Leu | Pro | Asn | Gly | Lys | Tyr | Gly | Val | Asp | Lys | Phe | Lys | 670 | 680 |
| TAT | CTA | TTC | TTA | AAG | ATT | AAA | CTA | AAA | GAT | TTA | ATG | CGT | GAT | GCT | GAT |
| Tyr | Leu | Phe | Leu | Lys | Asn | Lys | Leu | Lys | Asp | Leu | Met | Arg | Asp | Ala | Asp |
| 685 | 690 | 695 | 700 |
| TTT | GTC | CAA | CCT | CCA | TTA | ATT | TCT | ACT | TAC | TTT | AGA | ACT | TTA | TTG |
| Phe | Val | Gln | Pro | Pro | Leu | Thr | Tyr | Ile | Ser | Thr | Tyr | Phe | Arg | Thr | Leu |
| 705 | 710 | 715 |
| GAT | GCT | CCA | CCA | ACT | GAT | AAT | TAT | GAA | AAA | TAT | TTG | GTT | GAT | TCG | TCC |
| Asp | Ala | Pro | Thr | Asp | Asn | Tyr | Glu | Lys | Tyr | Leu | Val | Asp | Ser | Ser |
| 720 | 725 | 730 |
| GTA | CAA | TCA | CAA | GAT | GCT | TTA | CAA | CAA | GAT | GCT | AGT | GAT | TAA | CTA | AAT |
| Val | Gln | Ser | Gln | Asp | Val | Leu | Gln | Leu | Thr | Gln | Leu | Thr | Asn | Thr |
| 735 | 740 | 745 |
| ATT | GAT | ACT | AAT | GCT | AGA | GTT | GCA | TCA | AGT | GTT | ATT | GGA | TAT | GTT |
| Ile | Asp | Thr | Asn | Ala | Arg | Val | Ala | Arg | Val | Ser | Ser | Val | Ile | Gly | Tyr |
| 750 | 755 | 760 |
| GAA | CCA | TGC | ACA | TCA | GAA | CAT | AAA | ATT | GGT | TCA | GAA | GCA | TCG | TGT |
| Glu | Pro | Cys | Gly | Thr | Ser | Glu | His | Lys | Ile | Gly | Ser | Ala | Leu | Cys |
| 765 | 770 | 775 | 780 |
| AAA | ATG | GCT | AAA | GAA | GCA | TCT | AGA | TTA | GGA | AAT | CTA | GGT | TTA | GTA |
| Lys | Met | Ala | Lys | Ala | Ser | Arg | Leu | Gly | Asn | Leu | Gly | Leu | Val | Asn |
| 785 | 790 | 795 |
| CGT | ATT | AAT | GAA | GGT | AAT | TAC | AAC | AAA | TGT | AAT | AAA | TAT | GGT | TAT |
| Arg | Ile | Asn | Glu | Ser | Asn | Tyr | Asn | Lys | Cys | Asn | Lys | Tyr | Gty | Tyr |
| 800 | 805 | 810 |
| GGA | GTA | TAC | GAA | AAC | AAA | CTA | AAA | ACA | AAA | TAT | AGA | GAA | ATA |
| Gly | Val | Tyr | Lys | Gly | Val | Lys | Leu | Lys | Thr | Tyr | Arg | Glu | Ile |
| 815 | 820 | 825 |
| TTT | GAT | GTT | AAT | CCT | AAT | AAT | AAT | GAA | TTA | ATC | TCA | AGA | TAT | GTA |
| Phe | Asp | Cys | Asn | Pro | Asn | Asn | Asn | Gly | Leu | Ile | Ser | Arg | Tyr | Gya |
| 830 | 835 | 840 |
| TAT | AGA | ATA | ATG | GAT | TTA | CAT | AAA | ATT | GGA | AAA | ATT | TTG | GCA | AAT |
| Tyr | Arg | Ile | Met | Asp | Leu | His | Lys | Ile | Gly | Phe | Ala | Asn | Tyr |
| 845 | 850 | 855 | 860 |
| GAT | GAA | AGT | GAA | GAT | TCT | CCT | TGC | GAA | CAA | AGA | GTT | CAT | TAC | TTG |
| Asp | Glu | Ser | Glu | Ser | Pro | Cys | Glu | Arg | Arg | Cys | His | Tyr | Leu | Glu |
| 865 | 870 | 875 |
| AGA | GGT | CTT | TTA | TAT | GTT | CCT | GAA | TAT | GTA | CAT | CAC | AGA | TAT | CAA |
| Arg | Gly | Leu | Leu | Tyr | Gly | Pro | Glu | Val | His | Pro | Cys | Val | Thr | Arg |
| 880 | 885 | 890 |
| TCA | TGT | ACG | CCT | AAT | ACG | TTT | GGA | AAT | AAC | ACA | AAT | TGT | GTA | ACA |
| Ser | Cys | Thr | Pro | Asn | Thr | Phe | Gly | Asn | Asn | Thr | Asn | Cys | Val | Thr |
| 895 | 900 | 905 |
| AAT | GGT | GAA |CAA | CAC | GTA | TAC | GAA | AAT | AGT | GGT | GGA | GAT | AAT | GCA |
| Asn | Gly | Glu | His | Val | Tyr | Glu | Asn | Ser | Cys | Gly | Asp | Asn | Ala | Thr |
| 910 | 915 | 920 |
| TGT | GGA | AGA | ACA | TAC | AGA | TGT | GAA | AGA | AGA | AGT | GGT | GAA | TGG | AAT |
| Cys | Gly | Arg | Arg | Thr | Gly | Arg | Thr | Tyr | Arg | Ser | Arg | Phe | Glu | Thr |
| 925 | 930 | 935 | 940 |
| GAC | TAT | AGA | AAA | CCC | CAC | GTT | TAT | GAC | AAT | TGT | GCC | GAT | GCA | AAT |
| Asp | Tyr | Arg | Lys | Pro | His | Val | Tyr | Asp | Cys | Ala | Asp | Ala | Asn | Ser |
| 950 | 955 | 960 | 965 |

5,476,781
(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 464 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:2:


TGAATATTCT ACTTACAAAA TATGGATAAT CTTGTAATGT ATATTCAGTA TTGGTAAATAT
TATTTATAAT TTTATAGCC AAATGTTTGG TTTTATCAAG TGCATAATTT ATATGTTTAAC
CAACAATAGA ATCTAATGCA AACATTTAAG TATATATAAA AATAATTTTA TATTAACCTTA
AGATGTTCGA TTAATTTTAT GTCTGTTGATG TGAGTAAAAC ACACCAAGATA TGATATATAC
TATATCTTA ATTCCTCCTT TGAATCCTATG TCTATCAACCT GCAAATTTAT CACAGTATAA
TTTTCGAGTT TGCTTTGCA GATATAACCA AACATACATA ATGTTGGAGT TTTGTGGTTC
GGATGAAAAG CGTACTTTT

8138
8198
8258
8318
8378
8438
8457
<table>
<thead>
<tr>
<th></th>
<th>275</th>
<th>280</th>
<th>285</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lys Glu Asn Lys Lys Tyr Met Phe Ile Glu Ser Asp Met Thr Ile Lys</td>
<td>290</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td>Ser His Lys His Tyr Asn Ser Thr Pro Asn Thr Asn Tyr Ala Tyr Leu</td>
<td>305</td>
<td>310</td>
</tr>
<tr>
<td></td>
<td>Tyr Ile Asp Val Leu Ser Glu Tyr Leu Asn Asp Ile Phe Lys Asn Val</td>
<td>325</td>
<td>330</td>
</tr>
<tr>
<td></td>
<td>Asn Tyr Tyr Phe Phe Asn Thr Phe Glu Leu Gln Tyr Asp Ser Pro Asp</td>
<td>340</td>
<td>345</td>
</tr>
<tr>
<td></td>
<td>Cys Gly Met Phe Asn Ile Ile Phe Leu Tyr Tyr Ile Val Tyr Phe Asn</td>
<td>350</td>
<td>355</td>
</tr>
<tr>
<td></td>
<td>Ile Ser Lys Phe Glu Phe Lys Lys Leu Tyr Tyr Ser Met Ser Phe</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td></td>
<td>Ile Gly Asp Leu Leu Ala Ser Ser Tyr Arg Gly Ala Leu Phe Ile Ser</td>
<td>370</td>
<td>375</td>
</tr>
<tr>
<td></td>
<td>Arg Tyr Asp Ile Asn Ser Ile Asp Glu Phe Lys Asn Thr Leu Glu Ile</td>
<td>380</td>
<td>385</td>
</tr>
<tr>
<td></td>
<td>Phe Asn Ile Lys Lys Lys Phe Met Glu Leu Ile Asp Met Tyr Lys</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td></td>
<td>Lys Asn Ser Asn Arg Ile Met Asn Val Cys Ser Lys Ile Lys Asn Asp</td>
<td>400</td>
<td>405</td>
</tr>
<tr>
<td></td>
<td>Tyr Asp Ser Tyr Ile Asp Asp Glu Lys Asn Ser Leu Glu Ser Asn Ile</td>
<td>410</td>
<td>415</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 325 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:3:

<table>
<thead>
<tr>
<th></th>
<th>Met Phe Leu Val Tyr Phe Asn Thr Phe Leu Ile Ile Ile Leu Leu Phe</th>
<th>1</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gly Ile Ile Gly Ile Tyr Ile Leu Thr Phe Val Phe Asn Ile Asp Phe</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Leu Ile Asn Asn Asn Lys Ile Tyr Ile Leu Ser Tyr Asn Ala Thr Asn</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Ile Asn Asn Ile Asn Asn Leu Asn Leu Tyr Asp Tyr Ser Asp Ile Ile</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Phe Leu Thr Asn Phe Asn Ile Asn Asp Leu Leu Val Thr Glu Ala</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Asn Asn Leu Gln Asp Ile Pro Ile Phe Asn Val Asn Asn Ile Ser</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Asn Gln Tyr Asp Phe Tyr Ser Ala Ser Ser Ser Asn Asn Val Asn Ile Leu</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Asp Gln Tyr Leu Arg Lys Thr Leu Asn Ile Asn Arg Asn Phe Leu Leu</td>
<td>100</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Leu Arg Asn Thr Ser Leu Ala Ile Val Phe Asn Asn Asn Glu Thr Phe</td>
<td>115</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>Phe Arg Asn Thr Ser Leu Ala Ile Val Phe Asn Asn Asn Glu Thr Phe</td>
<td>135</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>His Cys Tyr Ile Ser Ser Asn Gln Asn Ser Asp Val Leu Asp Ile Val</td>
<td>145</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>Ser His Ile Glu Phe Met Lys Ser Arg Tyr Asn Lys Tyr Val Ile Ile</td>
<td>165</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>Ser His Ile Glu Phe Met Lys Ser Arg Tyr Asn Lys Tyr Val Ile Ile</td>
<td>175</td>
<td></td>
</tr>
</tbody>
</table>
Gly Glu Ile Pro Val Asn Asn Asn Ile Ser Ile Asn Asn Ile Leu Asn
1 85 1 85 1 90
Asn Phe Ala Ile Ile Thr Asn Val Arg Leu Ile Asp Lys Tyr Asn Ser
1 95 2 00 2 05
Ile Ile Ser Phe Leu Asn Ile Asn Val Gly Thr Leu Phe Val Ile Asn
2 10 2 15 2 20
Pro 2 25

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 78 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ser Ser Ser Lys Lys Asn Asn Leu Gly Tyr Phe Asn Asn Leu Lys
1 5 1 0 1 3
Thr Glu Gly Val Ser Gin Ser Gin Val Phe Lys Asp Asn Tyr Arg Pro
2 0 2 5 3 0
Gly Tyr Tyr Gly Leu Asp Thr Asn Ala Ala Asn Pro Ala Asp Val Tyr
3 5 4 0 4 5
Asn Thr Glu Ser Asn Lys Pro Ser Thr Val Asp Val Trp Gly Asp Lys
5 0 5 5 6 0
Arg Leu Glu Gly Lys Ile Ile Pro Lys Ser Lys Lys Lys Lys
6 5 7 0 7 5

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 161 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Ser Ile Phe Ile Tyr Tyr Ile Phe Asn Asn Arg Phe Tyr Ile Tyr
1 5 1 0 1 5
Lys Arg Met Asn Thr Val Gin Ile Leu Val Val Ile Leu Ile Thr Thr
2 0 2 5 3 0
Ala Leu Ser Phe Leu Val Phe Glu Leu Trp Tyr Tyr Ala Glu Asn Tyr
3 5 4 0 4 5
Glu Tyr Ile Leu Arg Tyr Asn Asp Thr Tyr Ser Asn Leu Glu Phe Ala
5 0 5 5 6 0
Arg Ser Ala Asn Ile Asn Phe Asp Asp Leu Thr Val Phe Asp Pro Asn
6 5 7 0 7 5 8 0
Asp Asn Val Phe Asn Val Glu Glu Lys Trp Arg Cys Ala Ser Thr Asn
8 5 9 0 9 5
Asn Asn Ile Phe Tyr Ala Val Ser Thr Phe Gly Phe Leu Ser Thr Glu
1 0 0 1 0 5 1 1 0
Ser Thr Gly Ile Asn Leu Thr Tyr Thr Asn Ser Arg Asp Cys Ile Ile
1 1 5 1 2 0 1 2 5
Asp Leu Phe Ser Arg Ile Ile Lys Ile Val Tyr Asp Pro Cys Thr Val
1 3 0 1 3 5 1 4 0
<table>
<thead>
<tr>
<th>Met</th>
<th>Ser</th>
<th>Asn</th>
<th>Val</th>
<th>Pro</th>
<th>Leu</th>
<th>Ala</th>
<th>Thr</th>
<th>Lys</th>
<th>Thr</th>
<th>Ile</th>
<th>Arg</th>
<th>Lys</th>
<th>Leu</th>
<th>Ser</th>
<th>Asn</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg</td>
<td>Lys</td>
<td>Tyr</td>
<td>Glu</td>
<td>Ile</td>
<td>Lys</td>
<td>Ile</td>
<td>Tyr</td>
<td>Lys</td>
<td>Lys</td>
<td>Asp</td>
<td>Gln</td>
<td>Asn</td>
<td>Thr</td>
<td>Cys</td>
<td>Phe</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Arg</td>
<td>Val</td>
<td>Val</td>
<td>Asp</td>
<td>Met</td>
<td>Val</td>
<td>Val</td>
<td>Pro</td>
<td>Leu</td>
<td>Tyr</td>
<td>Asp</td>
<td>Val</td>
<td>Cys</td>
<td>Asn</td>
<td>Glu</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Ser</td>
<td>Gly</td>
<td>Val</td>
<td>Thr</td>
<td>Leu</td>
<td>Gln</td>
<td>Ser</td>
<td>Cys</td>
<td>Ser</td>
<td>Pro</td>
<td>Asn</td>
<td>Ile</td>
<td>Gln</td>
<td>Val</td>
<td>Ile</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Asp</td>
<td>Asn</td>
<td>Thr</td>
<td>His</td>
<td>Val</td>
<td>Arg</td>
<td>Ile</td>
<td>Lys</td>
<td>Val</td>
<td>His</td>
<td>Gln</td>
<td>Asp</td>
<td>Thr</td>
<td>Leu</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>Lys</td>
<td>Glu</td>
<td>Met</td>
<td>Cys</td>
<td>Phe</td>
<td>Glu</td>
<td>Leu</td>
<td>Leu</td>
<td>Phe</td>
<td>Pro</td>
<td>Cys</td>
<td>Asn</td>
<td>Val</td>
<td>Asn</td>
<td>Gln</td>
<td>Ala</td>
<td>85</td>
<td>90</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Val</td>
<td>Trp</td>
<td>Lys</td>
<td>Tyr</td>
<td>Val</td>
<td>Ser</td>
<td>Arg</td>
<td>Leu</td>
<td>Leu</td>
<td>Leu</td>
<td>Asp</td>
<td>Val</td>
<td>Asn</td>
<td>Val</td>
<td>Ser</td>
<td>His</td>
<td>100</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>Asn</td>
<td>Asp</td>
<td>Val</td>
<td>Lys</td>
<td>Tyr</td>
<td>Lys</td>
<td>Leu</td>
<td>Ala</td>
<td>Asn</td>
<td>Phe</td>
<td>Arg</td>
<td>Leu</td>
<td>Thr</td>
<td>Leu</td>
<td>Asn</td>
<td>Gly</td>
<td>115</td>
<td>120</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>His</td>
<td>Leu</td>
<td>Lys</td>
<td>Leu</td>
<td>Lys</td>
<td>Glu</td>
<td>Ile</td>
<td>Asp</td>
<td>Gln</td>
<td>Pro</td>
<td>Leu</td>
<td>Phe</td>
<td>Ile</td>
<td>Tyr</td>
<td>Phe</td>
<td>130</td>
<td>135</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Asp</td>
<td>Asp</td>
<td>Leu</td>
<td>Gly</td>
<td>Asn</td>
<td>Tyr</td>
<td>Gly</td>
<td>Leu</td>
<td>Ile</td>
<td>Thr</td>
<td>Lys</td>
<td>Glu</td>
<td>Asn</td>
<td>Ile</td>
<td>Gly</td>
<td>145</td>
<td>150</td>
<td>155</td>
<td>160</td>
</tr>
<tr>
<td>Asn</td>
<td>Asn</td>
<td>Asn</td>
<td>Leu</td>
<td>Gln</td>
<td>Val</td>
<td>Asn</td>
<td>Lys</td>
<td>Asp</td>
<td>Ala</td>
<td>Ser</td>
<td>Phe</td>
<td>Ile</td>
<td>Thr</td>
<td>Ile</td>
<td>Phe</td>
<td>165</td>
<td>170</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Glu</td>
<td>Tyr</td>
<td>Ala</td>
<td>Tyr</td>
<td>Ile</td>
<td>Cys</td>
<td>Leu</td>
<td>Gly</td>
<td>Arg</td>
<td>Lys</td>
<td>Val</td>
<td>Tyr</td>
<td>Leu</td>
<td>Asn</td>
<td>Glu</td>
<td>180</td>
<td>185</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Val</td>
<td>Thr</td>
<td>Phe</td>
<td>Asp</td>
<td>Val</td>
<td>Thr</td>
<td>Thr</td>
<td>Asp</td>
<td>Ala</td>
<td>Thr</td>
<td>Asn</td>
<td>Ile</td>
<td>Thr</td>
<td>Leu</td>
<td>Asp</td>
<td>195</td>
<td>200</td>
<td>205</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Asn</td>
<td>Lys</td>
<td>Ser</td>
<td>Val</td>
<td>Asn</td>
<td>Ile</td>
<td>Ala</td>
<td>Val</td>
<td>Ser</td>
<td>Phe</td>
<td>Leu</td>
<td>Asp</td>
<td>Ile</td>
<td>Thr</td>
<td>Tyr</td>
<td>210</td>
<td>215</td>
<td>220</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Val</td>
<td>Asn</td>
<td>Asn</td>
<td>Gln</td>
<td>Lys</td>
<td>Asp</td>
<td>Leu</td>
<td>Lys</td>
<td>Leu</td>
<td>Lys</td>
<td>Lys</td>
<td>Leu</td>
<td>Lys</td>
<td>Arg</td>
<td>Gly</td>
<td>225</td>
<td>230</td>
<td>235</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Tyr</td>
<td>Gly</td>
<td>Glu</td>
<td>Phe</td>
<td>Glu</td>
<td>Val</td>
<td>Tyr</td>
<td>Asn</td>
<td>Ala</td>
<td>Asp</td>
<td>Thr</td>
<td>Gly</td>
<td>Lys</td>
<td>Gln</td>
<td>Ile</td>
<td>Tyr</td>
<td>245</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>Ala</td>
<td>Lys</td>
<td>Asn</td>
<td>Leu</td>
<td>Ser</td>
<td>Ile</td>
<td>Lys</td>
<td>Asn</td>
<td>Tyr</td>
<td>Asp</td>
<td>Thr</td>
<td>Val</td>
<td>Ile</td>
<td>Gln</td>
<td>Val</td>
<td>Glu</td>
<td>260</td>
<td>265</td>
<td>270</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Lcu</td>
<td>Pro</td>
<td>Val</td>
<td>Asn</td>
<td>Lys</td>
<td>Val</td>
<td>Arg</td>
<td>Ala</td>
<td>Tyr</td>
<td>Thr</td>
<td>Lys</td>
<td>Asp</td>
<td>Glu</td>
<td>Asn</td>
<td>275</td>
<td>280</td>
<td>285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Arg</td>
<td>Asn</td>
<td>Leu</td>
<td>Cys</td>
<td>Leu</td>
<td>Met</td>
<td>Lys</td>
<td>Ile</td>
<td>Thr</td>
<td>Ser</td>
<td>Ser</td>
<td>Thr</td>
<td>Gly</td>
<td>Val</td>
<td>Asp</td>
<td>290</td>
<td>295</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Glu</td>
<td>Tyr</td>
<td>Val</td>
<td>Thr</td>
<td>Ser</td>
<td>Asn</td>
<td>Asn</td>
<td>Ala</td>
<td>Lcu</td>
<td>Lys</td>
<td>Gln</td>
<td>Thr</td>
<td>Leu</td>
<td>Arg</td>
<td>Val</td>
<td>305</td>
<td>310</td>
<td>315</td>
<td>320</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Lys</td>
<td>Lys</td>
<td>Phe</td>
<td>Asp</td>
<td>Lys</td>
<td>Ser</td>
<td>His</td>
<td>Leu</td>
<td>Lys</td>
<td>Ile</td>
<td>Val</td>
<td>Met</td>
<td>His</td>
<td>Asn</td>
<td>Arg</td>
<td>325</td>
<td>330</td>
<td>335</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Ser</td>
<td>Gly</td>
<td>Asn</td>
<td>Val</td>
<td>Phe</td>
<td>Pro</td>
<td>Leu</td>
<td>Arg</td>
<td>Ser</td>
<td>Leu</td>
<td>Tyr</td>
<td>Leu</td>
<td>Glu</td>
<td>Leu</td>
<td>Ser</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Arg Val Ala Ser Ser Val Ile Gly Tyr Val Tyr Glu Pro Cys Gly</td>
<td>755 760 765</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr Ser Glu His Lys Ile Gly Ser Glu Ala Leu Cys Lys Met Ala Lys</td>
<td>770 775 780</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln Ala Ser Arg Leu Gly Asn Leu Gly Val Asn Arg Ile Asn Glu</td>
<td>785 790 795 800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser Asn Tyr Asn Lys Cys Asn Lys Tyr Gly Tyr Arg Gly Val Tyr Glu</td>
<td>805 810 815</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn Asn Lys Leu Lys Thr Lys Tyr Tyr Arg Glu Ile Phe Asp Cys Asn</td>
<td>820 825 830 835 840 845</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro Asn Asn Asn Asn Glu Leu Ile Ser Arg Tyr Gly Tyr Arg Ile Met</td>
<td>850 855 860 865 870 875 880</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Leu His Lys Ile Gly Glu Ile Phe Ala Asn Tyr Asp Glu Ser Glu</td>
<td>890 895 900 905 910 915 920 925</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser Pro Cys Glu Arg Arg Cys His Tyr Leu Glu Asp Arg Gly Leu Leu</td>
<td>930 935 940 945 950 955 960</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr Gly Pro Glu Tyr Val His His Arg Tyr Gin Glu Ser Cys Thr Pro</td>
<td>965 970 975 980 985 990 1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr Gln Asn Pro Ser Lys Cys Asp Ala Gly Cys</td>
<td>995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**(2) INFORMATION FOR SEQ ID NO.7:**

(i) **SEQUENCE CHARACTERISTICS:**

(A) LENGTH: 163 amino acids

(B) TYPE: amino acid

(C) TOPOLOGY: linear

(ii) **MOLECULE TYPE:** protein

(x) **SEQUENCE DESCRIPTION: SEQ ID NO.7:**

<p>| Arg Ser Ile Arg Leu Asn Ser His Lys Asp Leu Pro Gln Glu Tyr Arg | 1 5 10 15 |
| Tyr Val Asn Val His Phe Leu Ile Ser Tyr Thr Asn Arg Lys Ser | 20 25 30 |
| Val Asp Lys Glu Ile Leu Asp Ile Ile Lys Asp Lys Gin Gly Lys Ile | 35 40 45 |
| Asn Val Ile Phe Asp Leu Leu Lys Ser Ser Ser Ile Glu Ser Ile His | 50 55 60 |
| Asn Thr Tyr Lys Tyr Ile Glu Pro Ala Glu Asp Glu Ile Phe Asp | 65 70 75 80 |
| Thr Ile Arg Lys Thr Arg Met Lys Glu Met Asn Val Ser Asn Val Ile | 85 90 95 |
| Ile Asn Ile Lys Leu Tyr Pro Ile Ser Tyr Cys Lys Asp Tyr Asp Arg |</p>
<table>
<thead>
<tr>
<th>100</th>
<th>105</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>Thr</td>
<td>lle</td>
</tr>
<tr>
<td>115</td>
<td>120</td>
<td>125</td>
</tr>
<tr>
<td>Lys</td>
<td>Asp</td>
<td>Asn</td>
</tr>
<tr>
<td>130</td>
<td>135</td>
<td>140</td>
</tr>
<tr>
<td>Pro</td>
<td>lle</td>
<td>Phe</td>
</tr>
<tr>
<td>145</td>
<td>150</td>
<td>155</td>
</tr>
</tbody>
</table>

Tyr | Tyr | Glu |

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 1511 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDNESS: double
   (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) ORIGINAL SOURCE:
   (A) ORGANISM: Amsacta moorei orizopoulos virus

(iv) FEATURE:
   (A) NAME/KEY: CDS
   (B) LOCATION: complement (18.218)

(v) FEATURE:
   (A) NAME/KEY: CDS
   (B) LOCATION: complement (234.782)

(vi) FEATURE:
   (A) NAME/KEY: CDS
   (B) LOCATION: 82L.151

(vii) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
GAATTCAGT TAAATATTA TAAACAAACA TATATATTTTT TTAAGAATA TATAAAATT 60
TTTACATTAT TATATTATG TGTATAATTG TATATAATT GTTATTGTAT ATAAACATAA 120
TTATTTATCA TTTACCTAT TTTTTTTTT TATCTACTA ACGAAATACT AGATTGGGCA 180
CCTTCAATAT CAGAATATAT ATATATATTA TTTGGATATT ATGAATAATA ATATTAATAT 240
GAATTTATAT ACACACACAA ACGATATATT CTTGTTATTATT TATTTTCAAT 300
ATTTGATTATT TTATTTTTTT TTTAATGTAA AAGAAGACAT CTATATAACA AATTTGACAT 360
ATACCTCTGA ATTTTTTTTTTT TTATCTACTA TTAGGAATTATT ATTTGGATATT AAGATTAAAT 420
ATATTTCTAT TAAAGTCACA ATTTAATCAA GCAAACATAA CTTTTTTTT TATATTAGCA 480
ATTTATCAC AAAATTGTCT TAAATCTATT TCTCTAAAATA ATGACAGTC ATCTATGCCA 540
ATAATATCAT ATATATCTAC GATATGTAT ATCAATATTAT TTTATATT 600
AAATTTCTAT TATTTGCTAT ATTTGCTTAA TGATTTTTTT TATTTTATTAT 660
TTATCTATAT TATGATTATT AATACAGTAT TTTGGAATATT TAAATAATTAT 720
TTCCGATCTA ATTTCTGGGT TTTGCAAAAT AAGATAGGCA ATATTATAA TTCTATGCAC 780
ATTTTTTTTT TAAAATATTAA TAACTAATGAA AAAAAAATAA TATCAATT AGAAAATAAT 840
AAATATACAT AATGATTTAT TTAATGATATT TTAATGATATT 890
TTAAATTTATAT ATTTGCTTTTT TATATATT TATAAAATTATT 940
```

Translated:

```
Val | 87 |
89 |
93 |
98 |
```
<table>
<thead>
<tr>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATA ATA TAT TAT GAT AAT AAT ATT TTA AAT AAT ATT CCA GAA AAT ATT</td>
<td>1034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile Ile Tyr Tyr Asp Asn Asn Ile Leu Asn Asn Ile Pro Glu Asn Ile</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>AAA AGT TTA ATT TCA AAT TTA ATT ATT AAT TTA AAT TTT ATA</td>
<td>1082</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Ser Leu Tyr Ile Ser Asn Leu Asn Ile Leu Asn Phe Ile</td>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACA AAA TTA AAA AAT ATA ACA TAT TTA GAT ATA TCT TAT AAC AAA AAT</td>
<td>1130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr Lys Leu Lys Asn Ile Thr Tyr Leu Asp Ser Tyr Asn Lys Asn</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGC AAT ATA AGT AAT ATT ATA CTA CCA CAT TCT ATA GAA TTT TTA AAT</td>
<td>1178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser Asn Ile Ser Asn Ile Leu Pro His Ser Ile Glu Phe Leu Asn</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td>TGT GAA TCA TGT AAT ATA AAT GAC TAT AAT TTT ATT AAT AAT TTA GTA</td>
<td>1226</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys Glu Ser Cys Asn Ile Asn Asp Tyr Asn Phe Ile Asn Asn Leu Val</td>
<td>110</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>120</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>AAT TTA AAA TTA ATA ATA TCT AAA AAT AAA TTT GGT AAC TTT AAT</td>
<td>1274</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn Leu Lys Lys Leu Ile Ile Thr Tyr Asp Ser Ser Tyr Asn Lys Leu</td>
<td>130</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>135</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>AAT GTT TTT CCT ATT AGT GTA GGT GAG TTA AAT ATG GAA TCA ATA CAA</td>
<td>1322</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn Val Phe Pro Ile Ser Ile Val Glu Leu Met Glu Ser Ile Gin</td>
<td>145</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>155</td>
<td></td>
</tr>
<tr>
<td>ATA AAA GAT TAT AAA TTT ATA GAA AAA TTA ATT AAT TTA AAA AAA TTA</td>
<td>1370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile Lys Asp Tyr Leu Lys Asn Ile Leu Lys Lys Leu</td>
<td>160</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>165</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>GAT ATA TCT TTC AAT GTT AAA AAA AAT AAT ATA CAT TTT GAT AAT AAA TTT</td>
<td>1418</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Ile Ser Phe Asn Val Lys Asn Asn Ile His Leu Ile Lys Phe</td>
<td>175</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>180</td>
<td>185</td>
<td></td>
</tr>
<tr>
<td>CCA AAA AGT ATA ACT CAT TTA TGT GAT TAT CAA TCA TAT AAA GAA AAT</td>
<td>1466</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro Lys Ser Ile Thr His Leu Cys Asp Tyr Gin Ser Tyr Lys Glu Asn</td>
<td>190</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>195</td>
<td>200</td>
<td>205</td>
</tr>
<tr>
<td>TAT AAT TAT TTA AAA TTA TCA AAT ATA ATT GAA TAT GAA TTC</td>
<td>1511</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr Asn Tyr Leu Lys Asn Leu Ser Ser Ile Ile Glu Tyr Glu Phe</td>
<td>210</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>215</td>
<td>220</td>
<td></td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 66 amino acids
(B) TYPE: amino acid

(ii) MOLECULE TYPE: protein

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Glu Asn Asn Asp Asn Tyr Tyr Ser Asp Ile Glu Gly Ala Lys Ser
1 5 10 15
Asp Ile Ser Leu Val Asp Arg Lys Lys Ile Gly Lys Met Ile Asn
20 25 30
Asn Ile Val Asn Asn Asn Glu Leu Asn Lys Glu Leu Ser Asn Asn
35 40 45
Asn Lys Met Leu Lys Asn Leu Leu Asp Ser Leu Lys Lys Tyr Asp Cys
50 55 60
Cys Leu
65

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 182 amino acids
(B) TYPE: amino acid

1128
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met  Ser  Ile  Glu  Leu  Ile  Ile  Gly  Pro  Met  Phe  Ser  Gly  Lys  Thr  Thr
  1             5              10         15
Glu  Leu  Met  Arg  Lys  Ile  Asn  Arg  Tyr  Ile  Leu  Ser  Asn  Gln  Lys  Cys
   20            25          30
Val  Ile  Ile  Thr  His  Asn  Ile  Asp  Asn  Arg  Phe  Ile  Asn  Lys  Asn  Ile
   40            45
Ile  Asn  His  Asp  Gly  Asn  Ile  Leu  Asn  Lys  Glu  Tyr  Leu  Tyr  Ile  Lys
   50            55
Thr  Asn  Asn  Leu  Ile  Asn  Glu  Ile  Asn  Ile  Val  Asp  Asn  Tyr  Asp  Ile
   65            70         75         80
Ile  Gly  Ile  Asp  Glu  Cys  Gin  Phe  Phe  Glu  Asp  Leu  Glu  Gin
   85
Phc  Cys  Asp  Lys  Met  Ala  Asn  Asn  Lys  Lys  Val  Ile  Val  Ala  Gln
  100           105        110
Leu  Asn  Cys  Asp  Phe  Asn  Arg  Asn  Ile  Phe  Asn  Ser  Ile  Ser  Lys  Leu
  115           120        125
Ile  Pro  Lys  Val  Glu  Lys  Ile  Lys  Lys  Leu  Gln  Ala  Ile  Cys  Gin  Phe
  130           135        140
Cys  Tyr  Lys  Asp  Ala  Ser  Phe  Thr  Ile  Lys  Lys  His  Asn  Lys  Asn  Gin
  145           150      155        160
Ile  Ile  Glu  Ile  Gly  Glu  Gin  Asp  Leu  Tyr  Val  Pro  Val  Cys  Arg  Leu
  165          170        175
Cys  Tyr  Asn  Asn  Ser  Tyr
  180

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 220 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met  Asp  Leu  Leu  Asn  Ser  Asp  Ile  Ile  Leu  Ile  Asn  Ile  Leu  Lys  Tyr
  1             5              10         15
Tyr  Asn  Leu  Lys  Lys  Ile  Ile  Ile  Asn  Arg  Asp  Asn  Val  Ile  Asn  Ile
  20            25          30
Asn  Ile  Leu  Lys  Leu  Val  Asn  Leu  Glu  Glu  Leu  His  Ile  Ile  Tyr
   35            40
Tyr  Asp  Asn  Asn  Ile  Leu  Asn  Asn  Ile  Pro  Glu  Asn  Ile  Lys  Ser  Leu
   50            55          60
Tyr  Ile  Ser  Asn  Leu  Asn  Ile  Ile  Asn  Leu  Asn  Phe  Ile  Thr  Lys  Leu
   65            70         75         80
Lys  Asn  Ile  Thr  Tyr  Leu  Asp  Ile  Ser  Tyr  Asn  Lys  Asn  Ser  Asn  Ile
   85
Ser  Asn  Ile  Ile  Leu  Pro  His  Ser  Ile  Glu  Phe  Leu  Asn  Cys  Glu  Ser
  100           105       110
Cys  Asn  Ile  Asn  Asp  Tyr  Asn  Phe  Ile  Asn  Leu  Val  Asn  Leu  Lys
  115           120        125
Lys  Leu  Ile  Ile  Ser  Lys  Asn  Lys  Phe  Gly  Asn  Phe  Asn  Asn  Val  Phe
5,476,781 69 -continued

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAA GTNGATC CNGA ATATG

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

T TCAAA TTA ACTG CAACC

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGATGGATT TTAGAT TGC

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GCCTGTTGG GTAA CA CCTC

(2) INFORMATION FOR SEQ ID NO:16:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(iii) MOLECULE TYPE: DNA (genomic)

(xiv) SEQUENCE DESCRIPTION: SEQ ID NO:16:
CTGCTAGATT ATCTACTCCG

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(iii) MOLECULE TYPE: DNA (genomic)

(xiv) SEQUENCE DESCRIPTION: SEQ ID NO:17:
GTTCGAAACA AGTATTTCAT CTTTTAAAT AAATC

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(iii) MOLECULE TYPE: DNA (genomic)

(xiv) SEQUENCE DESCRIPTION: SEQ ID NO:18:
GAYGARGGRG GRCAATTYTT

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(iii) MOLECULE TYPE: DNA (genomic)

(xiv) SEQUENCE DESCRIPTION: SEQ ID NO:19:
GNCATGT TTYCNGG

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(iii) MOLECULE TYPE: DNA (genomic)

(xiv) SEQUENCE DESCRIPTION: SEQ ID NO:20:
GATGCAATAT CTAATTTTC

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3012 base pairs
(B) TYPE: nucleic acid
ATGAGTTAAG TACTTCGTTAC AACAACCAAA CATAGAAGAAT TATCAAATCG AAAATATGAA  60
ATGAGTTAAG TAGAAATGAG TTTTCGTAAT GTGATTAGAG TATGATTGTT  120
ATCCTATAGG ATGGGGTTGAA TGGAAAATCT TGGATTGATTAG TATGATTGTT  180
ATGAGTTAAG TTTGAAATTG CAAATCTCAT GATGGAATCA AAGT?CAGCG GCATTATTTA  240
AAAGAAATAG TGGTTGAGTT ATTGGTCGGG TGTGATGTAAC AGGAGGCCA AGTATGGAA  300
TAGATATTGC TAGTATTGCT AGATAATGCT TACACATAATG AGTAAATATG  360
ATTTTTTACT TGGACTTAAA TGGGAAACTAT TAAAAATCAA AAGGAAATCGA TGACCGCTTA  420
TTTTTTTTTT TTTGCAAGTC TTTGGGACTT TTTGACTTTA TAATAAGGAA AGATATCTAA  480
TATTTTTTTT TACAGTTAAA CAAAGATGCA TCAATTATTA CATATTATTC ACCATATTCG  540
TATTTTTTTT TAGTTAGAAA AGTATATTTA AATGGAAGAA TACCTTTTTG TGAATCTACA  600
GATGCAACTA TATATTTTTT AGATAATTTAT AATATCGGAT TATACGGACT ACTACCTTTT  660
GATATATAAT ACCAAGTTAA TAAAAATGAA CAAAATGTAT TATAAATTTA TTTATATTA  720
AGATACGGTG ATTTAGAAGGT CATAACGCAA GATACGGAGAT TAAATTATGC TAAAAATCTA  780
AGATAAAAAA ACAGGATAAA CAAATGCTCAA GTGAAAGATT TGCCCAATTT TTTGAAAAGT  840
AGAGCATATA CTAAGGATGA AATAGTCGCA AATCTATGTG TATGGAATAT AACATAGGTT  900
ACAGAAGGAG ACACCGGATT TCGGATTCTT ATTAAGTTAA TATGTTGATC GTCAGAGTA  960
TATAAAAAGT TTGATAAATT TTTTAAAATTT AATGTAATGC ATAACAAGGG AAGTGTTATA 1020
GTATTTTTTTC TAAATGCTTT ATATCGGAAA TTTGCTAATTG TAAAAAGGATA TCAGATTTAA 1080
GCATCTGATA CTTGCGGATT AGATTTTTGT TTTTCAATAT TAAAAATTTT TATATAGATA 1140
AACGACGAAA ATAAAAATTA AAGGGAAGAA ATGGAAGCGA AATAATAGTG CGGAAAGACA 1200
GTATCTCCAG ACAGTGTTAAA ATTTAGAATAT CTTAATCAA AAAAATCTCC CAAAATGCTCA 1260
TCTCTGAAAAC CAAAATAGTT TTTAAGCTTAA AGCTTTGTTGA GTATCGGCTG CTTTTTTTAA 1320
ACTCTTAGAT TATTTTTGTT ATGTGGATTA GATCTTGTG TATGAGTTAA TATTATTTTG 1380
CTTATGAAAAC ATGATATAGA GTGAGATTCT TTTGCTGATT TTTTATTTAA AGCTTTTGTT 1440
CTTATGAAAAC ATGATATAGA GGTGTTTTTGA GCTGTTAGTG TGAGTTGTTA 1500
ATATGAGAGT AGCGAATTAC TACACCAATA AGCAGATGAA AATATGATCA TTTTATTTTA 1560
ATATGAGAGT TTTGATTGAA TGGATATAAA ACTATATATG AGTATATCTA CTTTAAAAAC 1620
AGTAGAGAGA ACCCAATATA CATGCTATAT ATATTAGGAA TTTAACATAGG TATTTCTGTA 1680
AAATAGTATG TTGTGGCCGG TTTGGAAACA CCAAGACTCT CTTTTAAAAA CCCATACAGA 1740
TGTGCTAGAG TAGGAAATTG TTTAAATACT TTTTTGTTAT TTTTCAAATAA GAGATTTTAT 1800
GCTATGGCTA GAGACTCTGT AGTTTACTAA TTTAAGTGTA ACTATGTTAA TATGAAAAAT 1860
AACGAGCAG ATGCGGCAAT TCCCTGAGAT GCATATATAT TGGCAAGGAC TTTTGGTTTA 1920
ATTATGATT ATTATATAT TAAACCGAGTA GATAATCTAG CAGGATTACC AATTTTCCTT 1980
CCTATACATT TGGGAACTCA TCGGAAAAGCT TCTCAGGAAA AAAAATGATT AGGTGTGGTA 2040
AAGAAATTTA AATTCATTAT CTTAACAGGA AAAACCAAAAG ATTTAATTTG TGAGTCGTAT 2100
TTGGGCAATT TCTGCAATTTA TTTTTCTACT TACCTTTGAAA CTTTTTGAGA TGCTCCACA 2160
ACTGATAATTT ATGAAAAATAA TTGGTGTGAT TCCTGCCGCTA ATACACCTAAG GTCTTACGAT 2220
GGCTGTTGGATATACTTGGAA TACACCGTAA TACTATTGAT ACATAATTGCTA AAGTGTTGCTAC AAGGTGTTATT 2280
GGATGATTTT ATTGGACCTATCG GGAACATCTA GAACATAAATA TTGGTTCAGAA AGCATTTGT 2540
AAATGCGCTAT AAGACGGATCT TAGAGAACTAG AAACTGAGTGT ATGAAAATCTG ATGAAATGAA 2400
AGTAAATCACA ACAAATTGAAA TATAATATGCT TATAGAGGAG TATACGAAGAAA TAAACAACTA 2460
AAAAAAGAAAT TATATAGGAA AATATTGATG CTGAATCTCTA TATAAATATAAA AAAATATAATA 2520
TCCAGATATG GATATAAGATT ATAAGATTTTA CATAAAAAATG GAGAAATTTT TCAGAATTAC 2580
GATGAAAGGATG ATTCTCTCTG CGAAGCAAAGAG TGTATTACTAT GGGAAAGATT AGGGTTTTTATA 2640
TATGCTTCTG AAATGTACAA TCAGATATAT CAAGATCATG ATGACGCTA TACGTTGGA 2700
AATACACCAA ATGTTGTGAC AAGAGATGACT GAACACACCAG TATACGAAAAA TAGGGTGATGGA 2760
GATATGCCAA CAGTGGGAAG AAGACAGGTC TATGAGAACAA GAAGTAGGGA AGAAATGGAAT 2820
GACTATAGAAA AACCCTCAGT TTTATGACAT GTGGCGGCGA CAAATAGTCT ACTCTGAGAT 2880
AGCTGTTCACG AAGATGATGAT TAGTGTGGAA GCTGATACCTG ATTCAGATGTG ATGGTGGCGAC 2940
ACAGATGCTCA GTTTAGATTC TGATATAAGAA ATATGTATTC AAAATCTCCTAA AAAAAATGTG 3000
GCAGGATGCT AA 3012

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 419 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(x) MOLECULE TYPE: DNA (genomic)

TCAACTAATA AATAATATT TTATGCAATTT CAACATTTTGG GTATTTTTAAG TACAGAAGTT 60
ACTGGTATTT ATTTAACATA TACAATTTCT AGAGATTTGTA TTATAGATTT ATTTTCTAGA 120
ATTTATTAAA TAGTATATGAT TCCTCGTACT GTCGCAACATCT CAAACGATTAGA TATTATTATTA 180
AGATTATTTGA TGGGCAATAC ATCTATAAATA CATTATAAATA TATATTATAAT ATCAATGATA 240
ATTTTTATAT ATATTATTATC TAAAAGGACT TTTTTTTTTA TATAATAATAAATAAAATAAA 300
TGATGAACGT ACCTTTGGCA ACCAAATACAA TAAAGAAATAAT ATCAAATCGA AAAATGAAA 360
TAAAGATTTA TTAAAAGAGA GAAAATACCT GTTTCGAACAC TGTAGTACAT ATGTTGTATT 419

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 678 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(x) MOLECULE TYPE: DNA (genomic)

ATGGTTTCTAG TTTATTTAAA ACATATTTCAT TAAATAATTT TATATATTTGAT TATATAGGTT 60
ATTATATATG AAATTTTTGG GTTAATAATA GATTTTTTTAAA TAAATATAAT AAAAAAATATAT 120
ATTATATCAT ATACACCGAAC TAAATATAAC AATAAATAATA ATTTAATATT ATACGATTTAT 180
TCAAGATTTA TTTTTTGGAC AATATATTTAAC ATAAATATAAT ATCTTTTATGT AACAAAAGCT 240
AATAATTTAC AGATATACCC AATAATTTAT GAATAATAATA TATATATATAC TCAATATAAT 300
(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 486 base pairs (B) TYPE: nucleic acid (C) STRAND: double (D) TOPOLOGY: unknown

(iii) MOLECULE TYPE: DNA (genomic)

(xiv) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ATGTCATTAT TTTCTCTAAT TTTTTCCAAT ATAGATTTT ATATATATAA AAGAATGAAT 60
ACTGTACAAA TTATGTTGCT CATATATTA ACAAACAGCA TATCTTTTCT AGTTTTTCAA 120
TTAATGGATT ATGCCGAAAA TTATCAATAT ATATTTAAGAT ATATGATAC ATATCCTAA 180
TTACATTTTG CAGAGAGCCC AACATATAAT TTTGTAGATT TAACCTTTTT TGATCCCAAC 240
GATAAGTTTT TTTATGTTTG AGAAGAATGG CGCTGTGCTTT CAACTAAATA TAATATTTT 300
TATGCAGTTT CAACTTTTTGG ATTAAAAAGT ACAAGAAGTA CTGATTTAATTTATAACAT 360
ACAAATTCTTA GAGATTGGAT TATAGAATTA TTTTCTGAAA TTTAAAAATG ATATATGAT 420
CCTGTACTTG TCGAAAACAC TTTACATTTT TTTATGCTAC AGATTATAAAGATTAGTG AGGCCAATACA 480
TCATAA 486

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1395 base pairs (B) TYPE: nucleic acid (C) STRAND: double (D) TOPOLOGY: unknown

(iii) MOLECULE TYPE: DNA (genomic)

(xiv) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TAAATATTAG ATTCTACAC TTTCTCTTC ATATATGATAAAATATTATCATTATTATTTT 60
TATTTTACAT TACATCTTA TAATTCTATT ACTATTTTTT TTATACATAT CTATATATTC 120
CATAAACTTT TTATTTTTTT TTTAAATAT CGTCAATCTGTATTTTTAAATCGTCAATGAT 180
ATTAAATCA TATCTAGAAA TAAATATGAC ACCCTCTAATA CTACTAGCCA ATAAATCACC 240
AATAAAAATTC ATAGAATAAT ATATATTTTTT AAAATCAATA TTATGATTTT TGGAGAATAA 300
AACATATATA TATAAAAAAT TATATATATA CACAGCAGAA TGGAGAATTATAAT CAAATTTGA 360
TCCAAAGTTAT TAAATATTTAG AATATATCTAC TTTTTAAAT ATATATTTA AAATATCTGA 420
TAGTACATTTAC GTATATAATGT AATATATTCG TCTGAAATAAATTTT ATTTTATTGAT 480
GCTTTTTATAGTCAATACAT TATTTAATACATATAAATATTTTTTT TTTATGAGTTC 540
TGGTTATATA CACACTACTAT TAAAAAGTAA TGCACTTTTT TTATTTTTTCAAAGTTTT 600
(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 549 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:28:

| 60 |
| TTAATTAGAA TTTATTTAAC ATAATCTACA CACAGGAAACA TATAAAATCTT GTCCACCTAT |
| 120 |
| TTCAATTATT TGATTTTTAT TATGTTTTTT AATTGTAAPA GAAGCATCTT TATAACAAAA |
| 180 |
| TGACATATA GCTGTAATTT TTTTTTTTT TTTCTACTTT TTTATTTTTA TTTGATAGAA |
| 240 |
| ATTAATATA TTTCTGTTAA AGTCACAATTT TAATCCAGCA ACAATAAATT TTTTTTTATT |
| 300 |
| ATTAGGACAT TTTACCAAAA TTTGGTTCAA ATCTTTTTCT TCAAATAATT GACACTCACC |
| 360 |
| TATGCCAATA ATATCATAAT TATCTACGAT ATGGATTTCA TTAATTAAAT TATTTGTATT |
| 420 |
| ATGTATAAAA TTTCTTATT TTTAATATTT TCCGTCATGA TTTATTATAT TTTTTTTAT |
| 480 |
| AAATCTATTA TCTATATTAT GATTATAAT TACACATTTTG TATTAGATATA AAATATATCT |
| 540 |
| ATTAATTTTT CGCATCAATT CTGTTTTTTT GCCAGAAACA ATAGGACCAAT TATTAATTTCA |
| 549 |
| TATGACAT |

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:29:

| 60 |
| TTTTTTTAT TATTTGATAT ATTTTTTCAA AAAAAATTTA ATCAATGAAAA AAAAAATAAA |
| 69 |
| ATTATCAA |

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 141 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:30:

| 60 |
| AAACATAGGA CCAATTATTA ATCTCTACGA CTTTTTTTTT TATTATTGG TATATTATT |
| 120 |
| CAAAAAAAT TATATCAATG AAAAAAAAT AAAATTATCA AATGGATT TTACTAAATTCT |
| 141 |
| GATATATTT TATATAATAT T |

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown
(i) MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TTTAAAAA CAACTCATATT TTTTAAAAAG AATCTAAATAA AAAAAAACA ATTTTTATTT

ATTGGATAAT GTTTATTTTA AATCGTATT GATATTAAACA ATATTTTTA CATTTTACC

TATTTTTTTT TTTCTATCTA CTAAACGAAAT ATCAAGATTT GCACTTCCA TATCAGATAA

ATCAATATCA TTTTTTCGA T

(2) INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

(A) LENGTH: 660 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 32:

ATGGATTTAC TAAAAGCTGA TAAAAATTGA AAATAATTT TTTAAAAATA TAATTAAAAAA

AAAAAATAAA TAAAACAGAGA TAATGTATT AATTAATAA ATTAAAAAATA ATTAGTTAAT

TTAGAGAAT TGGATATAAT ATATATGAT AATAATTGT TAAATAATT TCCAGAAAT

ATAAAAAGTT TATATTTTCA AAAATTAATA TTTTTTTTAA AAAAAATTA

AAAATATAAA CATATTTGAG TATATCCTAT AAAACAAAAATA GCAATAAAGA TAATATTTA

CTAACACATT CTAAGAAATT TTTTATTTGA GATACTGTA ATATAAAGTA CATAATTTTT

ATTAATATT TAGTTATTTT AAAAAATAAT TAATATCTAA AAAATAATTG TGGTAACTTT

AATAAGTTTT TCTCTATTAG TATAGTGTAG TTTAATAGT CTTCAATTACA AATATAAGAT

TAAAATTTTT TGAAGAAAT TATTAATTAA AAAAAATAG ATATACCTT CAAAGTTAAA

AAAAATATAA TACATTGTGA AAAATTTCCA AAAATATCAA CTCAATTATG TGATTATCAA

TCATTAAAAAG AAAATTTAA TTTTTTAAA AAAAAATCTAA ATATATGTTG ATATGATTTC

(2) INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3907 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TTCTAAACC TTTATCTCCC CAAATCTCAG AGTAGATGTT TTTATTGATT CTGTTATTAAA

CACATCTGCT GGATTTGCCG CTTTGTATAC CAAACCATATA TATCCAGGTC TATAATTATC

TTAAAAACCT TGGGATTGAG ATACTCTTCT AGTTTTTTAA AAAAAATTTAT ATCCAAGATT

ATTTTTTTTT GATGAAACA TAATTGATA TATATATCTCT TATAGATATTG CAAATAATTAA

TTCTCATATT TTCAACATATT AGATTTTATTA TATAAAAAAG ATGAAATCT TATCAAAATT

TGGTTGCTAT ATTAAACAAC AAGCATATT CTTTTCTATG TTTCATTAT TAGTATTAG

CCGAAATTTA CGATATATTA TTAAGATATA ATGACATATA TTTCAATTAA CAAATTGCGA

GAAGCGCAA TATAAATTTT GATGATTAAA CGTTTTTTGA TCCCAGCAT AATGTTTTTA

ATGGCTAGAA AAAATGCGGC TGCTTCTCAA TTAATATAA TATATTATG GCAGTTTCAA

CTTTGGATT TTAAAGTACA GAAGATACTG ATTTTAATTT AACATATACA ATTCTAGAG

-continued
<table>
<thead>
<tr>
<th>TAGAACTTTA</th>
<th>TTGGATGCTC</th>
<th>CACCAACTGA</th>
<th>TAAATTATGAA</th>
<th>AAATATTTTG</th>
<th>TTGATTCGTC</th>
<th>3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTTACAAATCA</td>
<td>CAAGATGCTC</td>
<td>TACAGGGCTC</td>
<td>GTGGAATACCA</td>
<td>TCTAATACTA</td>
<td>TTGATACTAA</td>
<td>3060</td>
</tr>
<tr>
<td>TGCTAGAGTT</td>
<td>GCATCAAGTG</td>
<td>TATTTGGATA</td>
<td>TTGTATGAA</td>
<td>CCATGCGGAA</td>
<td>CATCAGAAC</td>
<td>3120</td>
</tr>
<tr>
<td>TAAATTTGGT</td>
<td>TCAAGAACAT</td>
<td>TGTGTAATAT</td>
<td>GCTAAAGAGA</td>
<td>GCATCTGATAG</td>
<td>TAGGAAATCT</td>
<td>3180</td>
</tr>
<tr>
<td>AGGTTTAGTA</td>
<td>AATCGTATTA</td>
<td>ATGAAAGTAA</td>
<td>TTCAACAAAA</td>
<td>TGTAATAATAAT</td>
<td>ATGGTTATAG</td>
<td>3240</td>
</tr>
<tr>
<td>AGGAAGTATAC</td>
<td>GAAATAAACA</td>
<td>AACCTAAAAC</td>
<td>AAAATATATT</td>
<td>AGAGAAATAAT</td>
<td>TTGATTGTAAA</td>
<td>3300</td>
</tr>
<tr>
<td>TCTTAATAAT</td>
<td>AATATGAAT</td>
<td>TATAATCCAG</td>
<td>ATATGGATAT</td>
<td>AGAATAATGG</td>
<td>ATTATACAA</td>
<td>3360</td>
</tr>
<tr>
<td>AATTGGAGAA</td>
<td>ATTTTTGCAA</td>
<td>ATTAGCAGTA</td>
<td>AAGTAGAATCT</td>
<td>CTTTGGCAAC</td>
<td>GAAAGTGTCA</td>
<td>3420</td>
</tr>
<tr>
<td>TTACTTGGAA</td>
<td>GATAAGACGT</td>
<td>TTTTATATGG</td>
<td>TCTGAATAT</td>
<td>GTAATACACAGA</td>
<td>GATATCAAAG</td>
<td>3480</td>
</tr>
<tr>
<td>ATCATGTACG</td>
<td>CTTATACGGT</td>
<td>TTGGAAATAAA</td>
<td>CACAAATTTG</td>
<td>GTAAACAGAA</td>
<td>ATGGTGAAAC</td>
<td>3540</td>
</tr>
<tr>
<td>ACACGTATAC</td>
<td>GAAATATGTT</td>
<td>GTGGAACATAA</td>
<td>TCCAAACATOT</td>
<td>GGAAAGAAAGA</td>
<td>CAGGATAGGG</td>
<td>3600</td>
</tr>
<tr>
<td>AAGAAAGAAGT</td>
<td>AGGATGAAT</td>
<td>GGAATGACTA</td>
<td>TAGAAAAACC</td>
<td>CACGTTTATG</td>
<td>ACAATTGTCG</td>
<td>3660</td>
</tr>
<tr>
<td>CGATGCAATAA</td>
<td>AGTCCCTTTT</td>
<td>CAGATAGCTG</td>
<td>TTCAACAGCAT</td>
<td>AGTATGAGTA</td>
<td>GTGAATCTGA</td>
<td>3720</td>
</tr>
<tr>
<td>ATCTGATTCG</td>
<td>GATGGAAGTT</td>
<td>GCGACACAGA</td>
<td>TGCTAATTGTA</td>
<td>GATTCGTATA</td>
<td>TTGAAAAATG</td>
<td>3780</td>
</tr>
<tr>
<td>TATCAAAAT</td>
<td>CCACTAAAAAT</td>
<td>GATGATGAGGG</td>
<td>ATGCTAATATG</td>
<td>AAAATATAATAAT</td>
<td>TTATATATAA</td>
<td>3840</td>
</tr>
<tr>
<td>TTAACCTAACG</td>
<td>ATGTAAAA</td>
<td>ATCAATAAAA</td>
<td>TGTATTTTTA</td>
<td>AAAATGATATT</td>
<td>ATGATGATTT</td>
<td>3900</td>
</tr>
<tr>
<td>GTGATAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3907</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:34:

| (i) SEQUENCE CHARACTERISTICS: |
| (A) LENGTH: 25 amino acids |
| (B) TYPE: amino acid |
| (D) TOPOLOGY: unknown |

| (ii) MOLECULE TYPE: protein |

| (ix) FEATURE: |
| (A) NAME/KEY: Region |
| (B) LOCATION: 3 |
| (D) OTHER INFORMATION: Note="This amino acid may be either Asn or Arg." |

| (ix) FEATURE: |
| (A) NAME/KEY: Region |
| (B) LOCATION: 12 |
| (D) OTHER INFORMATION: Note="This amino acid may be either Asn or Arg." |

| (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34: |

<table>
<thead>
<tr>
<th>Met</th>
<th>Ala</th>
<th>Xaa</th>
<th>Asp</th>
<th>Leu</th>
<th>Val</th>
<th>Ser</th>
<th>Leu</th>
<th>Leu</th>
<th>Phe</th>
<th>Met</th>
<th>Xaa</th>
<th>Xaa</th>
<th>Tyr</th>
<th>Val</th>
<th>Asn</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ile</th>
<th>Glu</th>
<th>Ile</th>
<th>Arg</th>
<th>Glu</th>
<th>Ala</th>
<th>Val</th>
<th>Xaa</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:35:

| (i) SEQUENCE CHARACTERISTICS: |
| (A) LENGTH: 17 amino acids |
| (B) TYPE: amino acid |
| (D) TOPOLOGY: unknown |

| (ii) MOLECULE TYPE: peptide |

<p>| (ix) FEATURE: |
| (A) NAME/KEY: Region |
| (B) LOCATION: 15 |
| (D) OTHER INFORMATION: Note=&quot;This amino acid may be either Thr or Ile.&quot; |</p>
<table>
<thead>
<tr>
<th><strong>SEQ ID NO:</strong></th>
<th><strong>SEQUENCE DESCRIPTION:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>Met Lys Ile Thr Ser Ser Thr Glu Val Asp Pro Glu Tyr Val Asn Ser</td>
</tr>
<tr>
<td>36</td>
<td>Asn Ala Leu Phe Phe Asn Val Phe</td>
</tr>
<tr>
<td>37</td>
<td>Glu Val Asp Pro Glu Tyr</td>
</tr>
<tr>
<td>38</td>
<td>ATGCGTACAG ATCTGTAAG TTTACTAT TTGTTAATAT TGAATTAAC</td>
</tr>
<tr>
<td>39</td>
<td>ATGAAAAATAA CAGTATACAG AGAATGAC CCCGACT ATACTAGTAA T</td>
</tr>
</tbody>
</table>

**INFORMATION FOR SEQUENCE DESCRIPTION:**

- **LENGTH:** 8 amino acids
- **TYPE:** amino acid
- **TOPOLOGY:** unknown
- **MOLECULE TYPE:** protein

- **LENGTH:** 6 amino acids
- **TYPE:** amino acid
- **TOPOLOGY:** unknown
- **MOLECULE TYPE:** peptide

- **LENGTH:** 66 base pairs
- **TYPE:** nucleic acid
- **STRANDNESS:** double
- **TOPOLOGY:** unknown

- **LENGTH:** 51 base pairs
- **TYPE:** nucleic acid
- **STRANDNESS:** double
- **TOPOLOGY:** unknown

- **LENGTH:** 24 base pairs
- **TYPE:** nucleic acid
- **STRANDNESS:** double
- **TOPOLOGY:** unknown
(1) MOLECULE TYPE: DNA (genomic)

(2) SEQUENCE DESCRIPTION: SEQ ID NO:40:

AATAAT AGAT TAATAA T G T A TT

(2) INFORMATION FOR SEQ ID NO:41:

(3) SEQUENCE CHARACTERISTICS:

A) LENGTH: 1689 base pairs

(1) MOLECULE TYPE: DNA (genomic)

(2) INFORMATION FOR SEQ ID NO:42:

TCCTATATT TGGTTTAATT CGATTCATT CCACCGCATA TCTAATATAA TTTATCAT 60
AAATACATTG AAATGATACG CTCGAGATCC AGCGTAAAGA AATAGCAAATA CTTTTACTTT 120
TTACATTAA TTTATYATCT AATATTTAT TCCGTTTAAT TCTATTCTCT TATTTTTA 180
AGTMTCTGA AAAAAATCAA TATAAGAATA ATTTAAATCA AAAAAACTA ATTTAAAACT 240
TGATATTCT CTAATATTTA AAAATGTTCC AATATTTAAT ACTTTTCTTC TCGAATTTAA 300
AATATTCTTA CAAGTTCTAT TATAATACCA GATAATTTGA TATATATTATA TATAATATT 360
TATATACCTG ATTGGAATAT TATTTTTAT TCTATATATTATA AATTCTTCAT 420
AAAAGATTCA GAAGAATTTAA TTTTTTTTGT AAATCTGAAGAA AATTTGAGAA GTTTTTCTTT 480
AAATATTAC TGTATATCTCA TATTATCTAA ATCTCTCTTTT ATTTATGAAT AAAAAAGC 540
TTTCTTGCAT ATTATATATAT ATTTTTTCTT GTTTATTTTTCTA AAAAACTAT 600
TATTATTTTTTTT TTTGCAATATCATC CATCTATATA ATTTGTTTCT GTTTAAACTAT CTGGCTCTAT 720
TAACCTTTTA TAAAGAAAATA AGCTAATATA TTGGTTCTTT AATTCTCTAA AAATAATTTA 780
CTTTCCATTAA TTATATATTT CTCTTTTTAT ATCTATAAAC TTTGTTCTAA GAAAACTAT 840
TTAATTATTAT AATTCCGAAT AATATTAGT TACTGGAGTA CGGCGACTAC ATAAAAATTT 900
ATTATTCTCG AATTTGCTAA ATTATTTAAA TTTTTATTAA ATAGGATAGTA AAAAACTTTTC 960
GATATTCTTTT TTATTTTTCAG TCTTGTGATAT TAAATTGGA AAAAACTTCTCA ATTTATTAG 1020
TAAACATCTT TTTATATATTA AAGAACATTTT ACAATGATTT AAAAAACTAT 1080
TAAACATA GTACGAATCAT AATAATATAA TTATATATTA CGATATTCTG ATATATATGA 1140
TCTTATAGTA TTATATTTTAA AATCTTTTAT TATTATTTTAT TTTATATATTA ATGTATTAT 1200
CCATCCTGGA AAATATTTCT TTTATATAT TATATATATAT ACAGCAGTTA ATGGTTTTCC 1260
CATACAGCTA TCTCACAAATA ACTAAATACCT ATCTCAATATT TTATATTTTAA TGAATATTCTT 1320
ACTTAAACAA TATGGAATAT CTTGTAATGT AATTTCAAGTA TTTGTAATAT TATTCAATAT 1380
TTATTTAGCC AAATGGTTGT TTTATCAAG TGCAATATTAT ATATGTTTAC CAACAATAAGA 1440
ACTTAATGCA AACATTTTATAT TATATATAAAAA ATATATTTTA AATATTTATA ATGTTTTCCA 1500
TAAATTTATTT GTCTGTGATT GGGATTTAAA ACCCAAGATA TGGATAATAC TATATACATTT 1560
ATTCTTCTTT TGAATCTACG TCTATCAATCT CCAAATTTTAT CCAATGATAAA TTTTCGAGTT 1620
TGTTTTGACG CATATAACCA AAACATACATA ATGGTGAGTT TGGTGTTTGCC GAGTAAAAAG 1680
CGTACTTTT 1689
( i ) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 485 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i i ) MOLECULE TYPE: protein

(x i ) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Phe Ala Leu Asp Ser Ile Val Gly Lys His Ile Asn Tyr Ala Leu</td>
<td>1 5</td>
</tr>
<tr>
<td>Asp Lys Thr Glu His Leu Pro Asn Lys Ile Met Asn Asn Ile Thr Asn</td>
<td>20 25</td>
</tr>
<tr>
<td>Thr Glu Ile Thr Leu Glu Asp Tyr Glu Tyr Phe Val Ser Arg Ile Phe</td>
<td>35 40</td>
</tr>
<tr>
<td>Ile Gly Leu Asn Leu Asn Ser Met Leu Leu Phe Trp Asp Thr Gly</td>
<td>50 55</td>
</tr>
<tr>
<td>Met Gly Lys Thr Leu Thr Ala Val Tyr Ile Ile Lys Tyr Ile Lys Glu</td>
<td>65 70 75 80</td>
</tr>
<tr>
<td>Leu Phe Pro Arg Trp Ile Ile Leu Ile Phe Ile Lys Lys Ser Leu Tyr</td>
<td>85 90 95</td>
</tr>
<tr>
<td>Ile Asp Pro Trp Leu Asn Thr Ile Arg Ser Tyr Ile Ser Ser Asp Thr Ser</td>
<td>100 105 110</td>
</tr>
<tr>
<td>Asn Ile Lys Phe Ile Tyr Tyr Asp Ser Ser Ser Ser Leu Asp Lys Phe</td>
<td>115 120 125</td>
</tr>
<tr>
<td>Asn Asn Ile Tyr Arg Ser Ile Glu Ser Ser Leu Asn Lys Lys Ser Arg</td>
<td>130 135 140</td>
</tr>
<tr>
<td>Leu Leu Ile Ile Ile Asp Glu Val His Lys Leu Ile Ser Arg Thr Val</td>
<td>145 150 155 160</td>
</tr>
<tr>
<td>Lys Lys Asp Asn Asn Glu Arg Asp Phe Thr Pro Ile Tyr Lys Lys Leu</td>
<td>170 175</td>
</tr>
<tr>
<td>Ile Lys Leu Ala Asp Phe Glu Asn Asn Lys Ile Leu Cys Met Ser Ala</td>
<td>180 185 190</td>
</tr>
<tr>
<td>Thr Pro Val Thr Asn Asn Ile Ser Glu Phe Asn Asn Leu Ile Gly Leu</td>
<td>195 200 205</td>
</tr>
<tr>
<td>Leu Arg Pro Asn Val Met Asn Ile Lys Glu Glu Tyr Ile Asn Asn Gly</td>
<td>210 215</td>
</tr>
<tr>
<td>Lys Leu Ile Asn Phe Lys Glu Leu Arg Glu Thr Leu Leu Ala Ile Cys</td>
<td>225 230 235 240</td>
</tr>
<tr>
<td>Ser Tyr Lys Arg Leu Ile Glu Ala Asp Ser Leu Thr Glu Thr Asn Tyr</td>
<td>245 250 255</td>
</tr>
<tr>
<td>Ile Asp Gly Tyr Ala Lys Lys Asn Ile Phe Tyr His Asn Ile Ile Met</td>
<td>260 265 270</td>
</tr>
<tr>
<td>Ser Asp Glu Glu Ser Lys Leu Tyr Asn Met Ala Glu Lys Tyr Asp Tyr</td>
<td>275 280 285</td>
</tr>
<tr>
<td>Lys Thr Glu Leu Gly Lys Thr Met Arg Arg Leu Ile Ser Ser</td>
<td>290 295 300</td>
</tr>
<tr>
<td>Phe Ala Phe Tyr Asp Leu Lys Ile Lys Gly Asp Leu Asp Asn Ile Glu</td>
<td>305 310 315 320</td>
</tr>
<tr>
<td>Tyr Asn Asp Met Ile Lys Arg Lys Leu Ala Glu Phe Ser Glu Phe Thr</td>
<td>325 330 335</td>
</tr>
<tr>
<td>Lys Asn Ile Asn Phe Ser Glu Ser Phe Ile Glu Ser Phe Lys Asn Asp</td>
<td>340 345 350</td>
</tr>
<tr>
<td>Asn Ile Lys Ile Lys Thr Asn Leu Pro Ile Thr Asp Ile Asn Asn Tyr</td>
<td>355 360 365</td>
</tr>
</tbody>
</table>

-continued
5,476,781

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 235 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:43:

AAAGTTTGA TAAATCATAC TTTAAAAATTG TTATGCATAAA TAGAGGAAGT GGTAAATGTAT 60
TCCCTATTAT AGACATATAT TGGAATTTA TGACCTGCAA AGGTTATCCT GTAAAAGCAT 120
CCGATACGTC TAGTTAGAT GTGTTTGTAT ATAAAATATGA TAAATAATAT ATGTTGAGA 180
AGAGAAGGAA AAAACCCGATTA TAGATGTGGA AGAGA 235

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 77 amino acids
(B) TYPE: amino acid
(C) STRANDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Lys Phe Asp Lys Ser His Leu Lys Asn Met His Asn Arg Gly Ser 1 5 10 15
Gly Asn Val Phe Pro Ile Arg Ser Leu Tyr Leu Glu Leu Leu Asn Val 20 25 30
Lys Gly Tyr Pro Val Lys Ala Ser Asp Thr Ser Arg Leu Asp Val Gly 35 40 45
Val Tyr Lys Leu Asn Lys Ile Tyr Ile Asp Asn Glu Asn Lys Ile 50 55 60
Ile Leu Glu Glu Ile Glu Thr Asp Tyr Arg Cys Gly Arg 65 70 75

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 235 base pairs

---Continued---
5,476,781

-continued

( B ) TYPE: nucleic acid
( C ) STRANDEDNESS: single
( D ) TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

AAAAGTTTGA TAAATCAGT TAAAAATCG TTATGCACAA TAGAGGAAAGC GGTAAATGTAT 60
TCCCTATTAG ATCACTATAT TTGGAAATTAT TGAACGTCAA AGGTATTACCT GTCAAAGCAGT 120
CGGATACGTC TAGTTAGAC GTTGGTGTGT ATAAACTAAA TAAATATAT ATGGATATAG 180
ATGAAAAATAA AATAATTTTA GAAGAAATCG AAACCGATTAG TACATGTTGA AGAGA 235

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 77 amino acids
( B ) TYPE: amino acid
( C ) STRANDEDNESS: single
( D ) TOPOLOGY: linear

(iii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Lys Phe Asp Lys Ser His Leu Lys Ile Val Met His Asn Arg Gly Ser
1  5  10  15
Gly Asn Val Phe Pro Ile Arg Ser Leu Tyr Leu Glu Leu Leu Asn Val
20  25  30
Lys Gly Tyr Pro Val Lys Ala Ser Asp Thr Ser Arg Leu Asp Val Gly
35  40  45
Val Tyr Lys Leu Asn Lys Ile Tyr Ile Asp Asn Arg Glu Asn Lys Ile
50  55  60
Ile Leu Glu Glu Ile Glu Thr Asp Tyr Arg Cys Gly Arg
65  70  75

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 10 amino acids
( B ) TYPE: amino acid
( C ) STRANDEDNESS: single
( D ) TOPOLOGY: linear

(iii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Lys Phe Lys Tyr Leu Phe Leu Lys Asn Lys
1  5  10

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 11 amino acids
( B ) TYPE: amino acid
( C ) STRANDEDNESS: single
( D ) TOPOLOGY: linear

(iii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Lys Ser Val Asn Ile Ala Val Ser Phe Leu Asp
1  5  10

(2) INFORMATION FOR SEQ ID NO:49:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(x) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Lys Tyr Leu Val Asp Ser Ser Val Gln Ser Gln

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:50:

AGATGATGAT TAAAGTGG

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GATAATGATA CTCCGTTGC

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GAAGTNGATC CNGAATATGT

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GAAAATAAAA TTAATTTGGA

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(xx) SEQUENCE DESCRIPTION: SEQ ID NO:54:
AGACAATTCG AGATATAATG

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(xx) SEQUENCE DESCRIPTION: SEQ ID NO:55:
CCGCATCTAT ATTCTGCTTC

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(xx) SEQUENCE DESCRIPTION: SEQ ID NO:56:
GTTAAAAACC TAAAAGTACCC

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(xx) SEQUENCE DESCRIPTION: SEQ ID NO:57:
TTTCAATTA AC TGCCAACC

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(xx) SEQUENCE DESCRIPTION: SEQ ID NO:58:
GGGATGGATT TTAGATTGCG

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO:59:

GTTGCATCTG TAGTTACATC

INFORMATION FOR SEQ ID NO:60:

SEQUENCE CHARACTERISTICS:
- LENGTH: 20 base pairs
- TYPE: nucleic acid
- STRANDEDNESS: single
- TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO:60:

TCTAGCAATA ATCGACTTAC

INFORMATION FOR SEQ ID NO:61:

SEQUENCE CHARACTERISTICS:
- LENGTH: 20 base pairs
- TYPE: nucleic acid
- STRANDEDNESS: single
- TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO:61:

GCCCTGGTGG GTAACAACTC

INFORMATION FOR SEQ ID NO:62:

SEQUENCE CHARACTERISTICS:
- LENGTH: 20 base pairs
- TYPE: nucleic acid
- STRANDEDNESS: single
- TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO:62:

CATTTCTATT AAGCCTAACG

INFORMATION FOR SEQ ID NO:63:

SEQUENCE CHARACTERISTICS:
- LENGTH: 20 base pairs
- TYPE: nucleic acid
- STRANDEDNESS: single
- TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO:63:

GTACCTTTA GCAAACCCAAAC

INFORMATION FOR SEQ ID NO:64:

SEQUENCE CHARACTERISTICS:
- LENGTH: 20 base pairs
- TYPE: nucleic acid
- STRANDEDNESS: single
- TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO:64:

CTGCTAGATT ATCTACTCCG
We claim:
1. A polynucleotide molecule free from association with nucleotide sequences with which it is associated in nature which comprises an Entomopoxvirus spheroidin polynucleotide sequence.
2. The polynucleotide molecule, according to claim 1, wherein said spheroidin polynucleotide sequence is from the Choristoneura fumiferana Entomopoxvirus.
3. The polynucleotide molecule, according to claim 1, wherein said spheroidin polynucleotide sequence is from the Choristoneura biennis Entomopoxvirus.
4. The polynucleotide molecule, according to claim 1, comprising the spheroidin polynucleotide sequence SEQ ID NO. 1.
5. The polynucleotide molecule, according to claim 1 comprising a polynucleotide sequence selected from the group consisting of a spheroidin gene coding sequence, a spheroidin gene regulatory sequence, and a spheroidin gene promoter sequence.
6. The polynucleotide molecule, according to claim 1, wherein said spheroidin sequence is further characterized by the ability to direct the expression of a heterologous gene to which said sequence is operably linked in a selected host cell or virus.
7. The polynucleotide molecule, according to claim 1, comprising said spheroidin polynucleotide sequence and further comprising a second polynucleotide sequence encoding a heterologous gene.
8. A recombinant molecule comprising a polynucleotide sequence encoding an Entomopoxvirus spheroidin promoter sequence, wherein said sequence is operably linked to a selected heterologous gene sequence, said sequence being capable of directing the expression of said gene in a selected host cell.
9. A recombinant polynucleotide molecule comprising an entomopoxvirus spheroidin promoter sequence operably linked to a heterologous gene, wherein said promoter is capable of directing the expression of said gene in a selected host cell capable of expressing said recombinant polynucleotide molecule wherein said promoter is from an entomopoxvirus selected from the group consisting of Amsacta moorei entomopoxvirus, Choristoneura fumiferana entomopoxvirus, and Choristoneura biennis entomopoxvirus.
10. A recombinant polynucleotide molecule comprising an entomopoxvirus Spheroidin promoter sequence operably linked to a heterologous gene, wherein said promoter is capable of directing the expression of said gene in a selected host cell capable of expressing said recombinant polynucleotide molecule wherein said promoter comprises SEQ ID NO. 22 which is capable of directing the expression of said heterologous gene.
11. A polynucleotide molecule free from association with nucleotide sequences with which it is associated in nature which comprises an Entomopoxvirus spheroidin polynucleotide sequence; provided that said Entomopoxvirus spheroidin polynucleotide sequence is homologous with an Entomopoxvirus spheroidin sequence which in nature encodes a spheroidin gene product having a molecular weight of greater than 50 kilodaltons.
12. A recombinant molecule comprising a polynucleotide sequence encoding an Entomopoxvirus spheroidin promoter sequence, wherein said sequence is operably linked to a selected heterologous gene sequence, said sequence being capable of directing the expression of said gene in a selected host cell; provided that said spheroidin promoter sequence is one which in nature regulates the expression of a spheroidin gene product having a molecular weight of greater than 50 kilodaltons.
13. A cell infected with a recombinant virus comprising an Entomopoxvirus spheroidin gene polynucleotide sequence operably linked to a selected heterologous gene sequence, provided that said Entomopoxvirus spheroidin gene polynucleotide sequence is one which is homologous to a gene which in nature encodes a spheroidin gene product having a molecular weight of greater than 50 kilodaltons.
14. A polynucleotide molecule free from association with nucleotide sequences with which it is associated in nature which comprises an Entomopoxvirus spheroidin polynucle-
107. A polynucleotide sequence provided that said Entomopoxvirus spheroidin polynucleotide sequence encodes the promoter or coding sequence of an Entomopoxvirus occlusion body protein.

15. A cell infected with a recombinant virus comprising an Entomopoxvirus spheroidin gene polynucleotide sequence operably linked to a selected heterologous gene sequence, provided that said Entomopoxvirus spheroidin gene polynucleotide sequence is one which in nature encodes the major protein of an Entomopoxvirus occlusion body.

108. An isolated polynucleotide molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:43, and SEQ ID NO:45.

17. An isolated polynucleotide molecule which encodes a gene product having the amino acid sequence of SEQ ID NO: 6.

* * * * *
UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,476,781
DATED : December 19, 1995
INVENTOR(S) : Richard W. Moyer, Richard L. Hall

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 7: Line 28: “fight” should read --right--.

Column 10: Line 4: “end Of the” should read --end of the--.

Column 13: Line 40: “techniques,” should read --techniques--.

Column 14: Line 60: “fights” should read --rights--.

Column 16: Line 18: “beating” should read --bearing--.

Column 18: Line 15: “β-interferon” should read --β-interferon--.

Column 19: Line 2: “hybrid. spheroidin-IFN-β gene” should read --hybrid spheroidin-IFN-β gene--.

Column 23: Line 65: “pro” should read --Pro--.

Column 26: Line 18: “(SEQ)” should read --(SEQ--; Line 63: “(SEQ D” should read --(SEQ ID--.

Column 26: Line 65: “to 465 1” should read --to 4651--.

Column 28: Line 53: “...A)GG (G/A)CA...TT” should read --...A)GG(G/A)CA...TT--.
UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 5,476,781
DATED : December 19, 1995
INVENTOR(S) : Richard W. Moyer, Richard L. Hall

It is certified that error appears in the above-indicated patent and that said Letters Patent is hereby corrected as shown below:

Column 29: Line 16: "(SEQ" should read --(SEQ ID--.

Column 29: Line 16: "(SEQ" should read --(SEQ ID--; "ORF O3" should read --ORF Q3--.

Column 30: Line 11: "vital" should read --viral--; Line 18: "grog" should read --grown--.

Signed and Sealed this Tenth Day of September, 1996

Attest:

BRUCE LEHMAN
Attesting Officer
Commissioner of Patents and Trademarks